This is a digital copy of a book that was preserved for generations on library shelves before it was carefully scanned by Google as part of a project to make the world's books discoverable online.

It has survived long enough for the copyright to expire and the book to enter the public domain. A public domain book is one that was never subject to copyright or whose legal copyright term has expired. Whether a book is in the public domain may vary country to country. Public domain books are our gateways to the past, representing a wealth of history, culture and knowledge that's often difficult to discover.

Marks, notations and other marginalia present in the original volume will appear in this file - a reminder of this book's long journey from the publisher to a library and finally to you.

Usage guidelines

Google is proud to partner with libraries to digitize public domain materials and make them widely accessible. Public domain books belong to the public and we are merely their custodians. Nevertheless, this work is expensive, so in order to keep providing this resource, we have taken steps to prevent abuse by commercial parties, including placing technical restrictions on automated querying.

We also ask that you:

+ *Make non-commercial use of the files* We designed Google Book Search for use by individuals, and we request that you use these files for personal, non-commercial purposes.

+ *Refrain from automated querying* Do not send automated queries of any sort to Google's system: If you are conducting research on machine translation, optical character recognition or other areas where access to a large amount of text is helpful, please contact us. We encourage the use of public domain materials for these purposes and may be able to help.

+ *Maintain attribution* The Google "watermark" you see on each file is essential for informing people about this project and helping them find additional materials through Google Book Search. Please do not remove it.

+ *Keep it legal* Whatever your use, remember that you are responsible for ensuring that what you are doing is legal. Do not assume that just because we believe a book is in the public domain for users in the United States, that the work is also in the public domain for users in other countries. Whether a book is still in copyright varies from country to country, and we can't offer guidance on whether any specific use of any specific book is allowed. Please do not assume that a book's appearance in Google Book Search means it can be used in any manner anywhere in the world. Copyright infringement liability can be quite severe.

About Google Book Search

Google's mission is to organize the world's information and to make it universally accessible and useful. Google Book Search helps readers discover the world's books while helping authors and publishers reach new audiences. You can search through the full text of this book on the web at `http://books.google.com/`

Stanford University Libraries
3 6105 025 685 848

STANFORD·VNIVERSITY·LIBRARY

ARCHIVES OF PSYCHOLOGY

EDITED BY
R. S. WOODWORTH

VOLUME VIII
1922

Columbia University

NEW YORK
1922

CONTENTS

Serial Number		Pages
53	Mood in Relation to Performance. Elizabeth T. Sullivan.	69
54	The Influence of Incentive and Punishment upon Reaction Time. Albert M. Johanson.	54
55	Psychological Tests Applied to Factory Workers. Emily T. Burr.	93
56	A Study of the Relation of Accuracy to Speed. Henry E. Garrett.	104
57	An Experimental Study of Hunger in Its Relation to Activity. Tomi Wada.	65
58	Individual Differences as Affected by Practice. Georgina S. Gates.	74

MOOD IN RELATION TO PERFORMANCE

BY

ELIZABETH T. SULLIVAN, Ph.D.

ARCHIVES OF PSYCHOLOGY

EDITED BY R. S. WOODWORTH

No. 53

NEW YORK
MAY, 1922

AGENTS: G. E. STECHERT & CO. : London (2 Star Yard, Carey St., W. C.); Paris (16, rue de Condé)

TABLE OF CONTENTS

Introduction. A Brief Summary of Some Definite Contributions to the Literature. 5

Chapter I. The Problem, Method and Results. A General Statement. 16

Chapter II. Test and Methods of Handling Data. 22
Section 1. Description of the Tests Used.
Section 2. Method of Handling Scores for the Objective Rating of Mood.
Section 3. Method of Eliminating Practice Effect from the Original Scores.

Chapter III. Discussion of Results. 37
Section 1. The Relation of Mood to Performance in the Various Tests.
Section 2. Combination of Tests to Secure a Measure Freer from Attenuation.

Chapter IV. Supplementary Studies. 56
Section 1. Circulation and Performance.
Section 2. Mood and the Weather.

Bibliography. 70

THIS research was undertaken in the psychological laboratory of Columbia University. To Professor Woodworth I owe a debt of gratitude for encouragement and stimulating criticism. I am also conscious of what I owe to Professor Poffenberger and Professor Hollingworth for valuable suggestions and to the graduate students in the departments of Psychology in Columbia University who acted as subjects and judges in the experiment.

Acknowledgments are due to Dr. Marcus B. Heyman, Head of the Manhattan State Hospital, who made it possible to extend the study to include patients at the Hospital. To the members of the Medical Staff, nurses, and attendants and to the patients thanks are especially given for generous co-operation.

Mood in Relation to Performance

INTRODUCTION

Brief Summary of Some Definite Contributions to the Literature

Temperament is a general term for the characteristic emotional state of individuals. The traditional classification of temperaments is into the sanguine, melancholic, choleric and phlegmatic. The sanguine and the choleric are supposed to be subject to rapid alternation of interest. With the choleric, interest is intense while it lasts whereas with the sanguine it is relatively feeble. The interest of the melancholic and of the phlegmatic is persistent when once it is aroused, though they are not easily excited. The interest of the melancholic is intense whereas that of the phlegmatic is lacking in vividness.[1]

The ancient and medieval schools of medicine were dominated by two theories, one that human lives were influenced by the planets and the other that all diseases were due to the excess or defect of one of the four humors—yellow bile or choler, black bile, blood and phlegm.[81] These physicians have left records that showed them to be shrewd observers, the evidence of recent times going far to confirm their observations that mental qualities do tend to occur in groups, the total mental constitution of a man being termed his temperament. Though a temperament is not always present in its purity or in a high degree, yet where it is present, the person who possesses it presents a likeness in mental characteristics to another such as is rarely presented by the different members of a single family.

According to Davenport[8] temperament is inherited, periodic outbursts of violent temper being a Mendelian dominant. He has worked out formulae for expressing the possibilities in the inheritance of choleric, cheerful, phlegmatic, melancholic and nervous temperaments and their various combinations.

A. In recent years much prominence has been given to the part played by the bodily states in producing exciting emotion.

The evident connection between the visceral disturbances and the psychic states is admitted by practically all but there is division of opinion as to the manner of connection. This suggests at once the theory of James,[15] which is that the bodily changes follow directly the perception of the exciting fact and that the feeling of these same changes as they occur *is* the emotion. Lange[16] and Sergi[22] hold that the basis of *all* feeling and emotion is physiological, visceral and organic and is located in the vasomotor system.

Sherrington[23] made some tests of this view by experiment. He chose for his subject a "markedly temperamental" dog and made spinal and vagal transections which cut off completely and immediately the sensation of the viscera and of the skin and muscles beyond the shoulder. The procedure at the same time cut from connection with consciousness the whole of the circulatory apparatus of the body. As a result of careful experimenting on the dog in this condition he concludes: "We may with James accept visceral and organic sensations and the memories and associations of them as contributory to primitive emotions but we must regard them as reinforcing rather than initiating the psychosis. Organic and vascular reaction, though not the actual excitant of emotion, strengthens it." He further states that we are forced back toward the likelihood that visceral expression of emotion is secondary to the central action occurring with the psychical state.

Cannon[3] has conducted a series of investigations in the Harvard Physiological Laboratory concerned with the bodily changes which occur in conjunction with pain, hunger, and the major emotions—fear and rage. A group of remarkable changes in bodily economy has been discovered all of which indicate that the bodily changes are responses that are adapted to the individual's welfare and preservation. Digestion is affected to an important degree by feelings. Pleasant taste and smell and sight of food clearly stimulate the digestive processes to action. On the other hand pain and great emotional excitement interfere seriously with the starting of the process or its continuation after it has been started. These effects extend to the entire alimentary canal, effectively stopping the gastric and intestinal peristaltic action both in man and in lower animals. That the mere sight or smell of favorite food may start the pouring out of gastric juice was noted many years ago by Bidder[2] and Schmidt[5] in a hungry dog

which had a fistulous opening through the bodily wall of the stomach. The observation, reported in 1852, was confirmed later by Schiff and also later by Pavlov[8] of Petrograd through admirably worked out experiments on dogs.

Other experiments have been undertaken to determine the part played by adrenalin since the action of the sympathetic division of the nervous system under pain, fear and rage is similar to that produced by adrenalin. When adrenalin is injected into the blood it will cause the pupils to dilate, the hair to stand erect, the blood vessels to be constricted, the action of the alimentary canal to be inhibited and sugar to be liberated from the liver. According to Cannon, "every one of the visceral changes noted—the cessation of processes in the alimentary canal (thus freeing the energy supply for other parts); the shifting of blood from the abdominal organs, whose activities are deferable, to the organs immediately essential to muscular exertion (the lungs, the heart, the central nervous system); the increased vigor of contraction of the heart; the quick abolition of the effects of muscular fatigue; the mobilizing of energy-giving sugar in the circulation—every one of these visceral changes is directly serviceable in making the organism more effective in the violent display of energy which fear or rage or pain may involve."[6]

Following Darwin's conclusions, Crile[7] makes the assumption that the discharge of nervous energy is accomplished by the laws of inheritance and association. He maintains that "man is a unified mechanism responding in every part to the adequate stimulus given it from without by the environment of the present and from within by the environment of the past the record of which is stored in part in cells throughout the mechanism but especially in the central battery—the brain." On the basis of the mechanistic theory he explains emotions as "ractivations of the entire motor mechanism for fighting, for escaping or for preserving the species." The significant thing here is not so much the theory as the actual observations on which the theory is based. As to the actual physiological results of emotion both Crile and Cannon agree that in fear all the motor mechanism is correlated for combat, the non-combatant organs being inhibited.

Closely connected with temperament and personality are the internal secretions[10] and the ductless glands[16] which give rise to them. Investigations in this field are strengthening the

opinion that the human organism is a "moving chemical mixture" dominated by the action and interaction of eight or ten ductless glands. The personality or temperament is thought to be due largely to the action of these glands and the degree of balance established between them. When for any reason the balance is disturbed between the system of glands disease is the result. Although the investigations of the endocrinologists are becoming increasingly prominent more need not be said about them here beyond this brief recognition of the importance of their findings to any study directly or remotely connected with temperament and personality.

B. The remaining works to be reviewed have been selected from the literature dealing with the investigation of the personality because of the definite nature of the study and the value of the method employed. Ach's work on Will[1] and Temperament, apart from the method used, is important because of the problems it presents for future development. The object of this study was to find a measure of the strength of will, the efficiency of the will being the relation of the actual results to the intended results. In any test the results might be missing, they might be changed in form or delayed. He used in combination the Saving Method, Repetition, and Paired Associates to present definite tasks to the subjects. Inner obstacles in the nature of previously formed habits were set up to interfere with the carrying out of the undertaking. These obstacles were of different strength and by the methods used the strength of the various habits could be determined. By the formation of habits of different strengths which would act as obstacles, Ach believed he could bring out weak and strong impulses, efficiency (besides other factors) depending on the strength of the determining tendency.

The work of Webb on Character and Intelligence represents an attempt at an exact study of character. The author's holds that there are two aspects to intelligence, one a general factor of mental energy, G—some general factor of the cortex— on which all performances depend, and a second factor W, prominent on the character side of mental activity as distinguished from the purely "intellective" side. He includes under intelligence all those qualities which are especially related to mental performances. Under character he includes volitional social and moral qualities. His subjects were 98 second year students at a training college in London, average age being

21 years, a similar group of 96 other students, who constituted the class the following year, and 4 groups of school boys about 12 years old. His method was that of group judgment, the judges being 20 prefects and 4 class-masters of the school. Each judge estimated each subject as to certain intellectual and character qualities, two judgments being secured from each judge on each quality to obtain reliability. The marks were given under conditions controlled to shut out any bias. The subjects were ignorant of the fact that the experiment was being carried on. The judgments were given with regard to the individual as a whole. On the basis of the estimates thus secured the investigator concludes: "We may say that the possession of a good deal of G, i.e., of pure intellectual ability (1) is revealed in most attempts to estimate intelligence qualities and (2) tends to occur in persons with stability of emotions, some cheerfulness added to a fair degree of sociability, with marked application to duty and some foresight and perseverance." The investigator attempts to show statistically that the factor G is functionally operative in producing correlation between mental performances, the correlation being always partly due to the G and partly due to specific abilities called into play by the specific nature of the performance, i.e., to the existence of a second general factor, W. on the side of character. He made a study of these character qualities using precisely the same method he used for the purely intellectual factor. He found that "stable qualities" of character correlated more highly with "profoundness estimates" and that the "unstable qualities" correlated best with "quickness." The nature of the second general factor W is closely related to "persistence of motives" and according to Webb is not unlike Ach's 'will.' He further states that "the theory of the second factor, persistence of motives, seems to be in good accordance with the system of relations which are exhibited by the two aspects of intelligence" for "profound intelligence, as distinguished from mere quickness may be regarded as being a steadier or more stable grasp of the mental content.' The investigation represents a definite attempt to study character in connection with intelligence and suggests a method of work that is of more than ordinary value.

Studies on judgment have furnished methods that are important if not indispensable in any investigation of character or temperament. Cattell's[4] study of American men of science

was an attempt to find a relation between the character of an individual and the way he expresses himself. He had ten judges for each science to judge the men in each branch of science. The judges were all prominent in the scientific field. The judgments were arranged in the order of merit, the investigator taking the P. E. as measuring the validity of the various judgments. If any man's average position varied from that of another by three times the P. E. there was no chance that the two men's places would be interchanged.

Hollingworth[11] made an experimental study of the characteristics of judgments of personal efficiency and the relation between judgment of one's own performance and judgment of the work of another person. He found that "judgments of one's own work tend to be only slightly better, from the point of view of correctness, than judgments of the work of another person." From the results of "An Experimental Study of Self Analysis," in which the same author collaborated with Cogan and Conklin, it appears that in general "the error of self-estimation tends to be half again as great as the average error of the judgments of associates," and in the long run a person who knows himself best is the best judge of others, or at least, tends to be a good judge (varying with the trait in question). Furthermore, there seems to exist a considerable degree of correlation between academic record and mental tests and the ability to judge.

A very recent study by Rugg emphasizes certain conditions that are essential to validity in judgment ratings. The independent judgments of three judges are necessary; the rating scales must be comparable and equivalent and the judges must be competent to rate. In the opinion of the author even with these three provisions judging methods are subject to a large error. In his own scale all three provisions are met. While he inclines to the belief that all rating scales are only relative he thinks that it is better to use the objective method of measuring than to attempt to improve the subjective methods in use.

A recent addition to tests of personality is the Pressey Crossing-out Test[20] devised for the investigation of emotional interests and distractibility. A part of the test is designed as a "measure of affective spread or tendency to emotionalize." Another is designed for measuring emotional distractibility, the test consisting of crossing out irrelevant words "sown"

into two paragraphs, one of commonplace matter and the other of matter calculated to excite emotion. The idea being that the emotional excitement of the second paragraph would cause the subject to overlook more irrelevant words, the difference between the scores of the two paragraphs becomes the measure of "emotional distractibility." A third part of this group scale is constructed to measure "moral discrimination and experience" and is derived directly from the Kent-Rosanoff tests for association in insanity. A fifth part of the scale is intended to measure"emotional memory" two lists of words being given to test memory for emotional words compared with that for unemotional words. The test as a whole is designed to meet the conditions required in a group test, the "crossing out" feature making it easy to give and to score. Thus far the literature reviewed has dealt with the subject of emotion in a general way, and with studies that have furnished something in the way of method bearing more or less directly on the treatment of the data used in the study with which the writer is concerned.

C. The immediate predecessors, however, to this study of mood and performance are some recent investigations in which performance is studied in relation to other conditions. The investigation made by Dr. L. S. Hollingworth[14] of Functional Periodicity and its effect on the mental and motor abilities of women is one of these. It covers the traits of voluntary speed of movement, steadiness, speed and accuracy of perception and controlled association all of which were tested under controlled experimental conditions. Contrary to general expectation, the results did not reveal any mental or motor inefficiency in connection with physiological periodicity.

The condition of ventilation in relation to mental work was studied by Thorndike, McCall and Chapman.[27] The object of this extended research was to investigate the effect of certain atmospheric conditions on efficiency. The subjects were tested in various functions under varying external conditions of temperature and humidity and air circulation, the experiment being controlled to eliminate practice and individual difference. The results of the experiment with daily change of air conditions showed that "when a person exerts himself his achievement at mental work is as good under various bad conditions as under the best conditions" (within the limits of the experiment). The part of the experiment having to do

with the rate of improvement made under prolonged bad and good conditions showed that the "improvement made during the week of a bad condition was equal to the improvement made during a best condition week." A third part of the experiment was concerned with the effect of the condition of air on mental work when the "instruction of the experiment, the conduct of the experiment and the nature of the tests were such as not to stimulate effort but to encourage carelessness." Work was done under conditions of bad air, and made "intentionally drearily monotonous and barren of interest," but still showed no inferiority in quality to that done under contrasting conditions.

Dr. Lorle I. Stecher made a similar study on the "Effect of Humidity on Mental Efficiency." This extended experiment was conducted under equally controlled conditions. The results as in the case with ventilation were negative indicating that "behavior or product per unit of time is not affected to any extent by varying degrees of humidity." Another study by Thorndike[28] on fatigue in school children is of especial interest here. The purpose of this study was to test the ability to do mental work before and after periods of hard mental exercise. The greater number of "after-work" tests were undergone when the subjects, if left to themselves, would not have engaged in mental work. Professor Thorndike says, "The children in the schools where I gave tests early and late in the day agreed that they were tired in the late hours and thought that they could not work nearly so well yet these same children *did* do just as well in the tests given. The relation between the feeling and the fact of fatigue then is not a simple relation of direct dependence but a very involved relation or series of relations." He concludes that "we can *feel* mentally fatigued *without being* so. Feelings of being 'mentally tired' serve as signs to us to stop working before our actual ability to work has suffered any important decrease."

The three studies just reviewed represent efforts made to study performance under conditions that involve a temporary disinclination to work. The remaining studies represent efforts made to study performance in relation to a diseased condition of some part of the organism.

Wallin[29] in "Mental Health of the School Child," reports the results of a study made on a group of twenty-seven public school children in Marion School, Cleveland, Ohio, all of whom

were handicapped to a considerable degree by diseased teeth or gums or an otherwise unsanitary oral cavity. The experiment extended over one year, corrective treatment of a complete order having been given during the time. Five series of psychological tests, given at stated intervals during the course of the treatment, served as a measure of the effect of dental treatment upon the working efficiency of the pupils. According to this investigator, there was a decided gain in every test and not only were the "gains more frequent than the losses" but the largest gains were invariably "emphatically larger than the largest losses." There was corroborative evidence to show that there was a general "improvement in the mental efficiency of these pupils." Most of the members of this experimental groups were "laggards and repeaters pedagogically retarded in their work from one to four years but during the experiment only one failed of promotion while six did thirty-eight weeks of work in twenty-four weeks and one boy finished two years of work within the experimental year." It is to be noted that no control group was used in this investigation.

A very recent study by A. H. MacPhail shows the effect on school work due to the removal of diseased tonsils and of adenoids. As the result of carefully controlled observations made on thirty-one pupils of the Adams and Cranch Schools, this investigator concludes that the "removal of diseased tonsils and adenoids is a factor in beneficially influencing the mental life of the school child. Not only is the health impaired by failure to remove these diseased parts but mental life and activity of the child as well."

A similar investigation with an additional purpose in view was made by Dr. Margaret Cobb,[5] one part of the work being concerned with the effect of the removal of tonsils and adenoids on efficiency and the other with the effect on the Intelligence Quotient. An extended experiment was carried on, the subjects of which were pupils in Public School 64, Borough of Manhattan, New York City, or patients at the Manhattan Eye, Ear and Throat Hospital. The hypothesis was that if a "child's working efficiency is lowered by the effect of adenoids and bad tonsils then removal should, unless such lowering be permanent, be followed after a reasonable time by an improvement." Care was taken to compare the improvement in efficiency with the change in efficiency of a control group whose

members had not been treated. Observations were made during a period of twelve months following the operation for adenoids and tonsils. Test results for the first six months were compared with those for the last six months and may be summarized as follows: In weight and height there was a slight but not a reliable increase during the first period and no reliable increase in height-age; there was no gain in strength of grip (possibly a decrease) and in speed of tapping. Fatigability was not lessened. During the second period, the results showed a noticeable improvement in weight, height-age and speed of tapping; there was no increase in height, no decrease in strength of grip and no lessening in fatigability.

In addition to the more extended experiment a comparison was made as to the effect of removal of tonsils on the intelligence levels of two groups of children, the one "selected for the presence of tonsils and the other for freedom from them," both groups being selected from Public School 64. There was no rise in the I. Q. as a result of the operation and a group of 236 children with diseased tonsils showed equal distribution of I. Q, with a group of 294 children who were normal in this respect.

Important among this class of investigations on the relation of disease to efficiency is that pursued by the Rockefeller Foundation under the direction of Dr. E. K. Strong[25] with the purpose of examining the effect of hookworm disease on the mental and physical development of children. The children were divided into five groups and tested at intervals of three and one-half months by use of the following tests: Opposites, calculation, logical memory, memory span, handwriting and the Binet-Simon Tests. Before any experimenting was done the children were divided into sub-groups on the basis of performance in a test given for the purpose of obtaining sub-groups that were nearly equal so that improvement could be compared between equal sub-groups. The improvement made by the uninfected group was taken as 100 per cent. By comparison the group "infected but not treated" showed the least improvement, 34 per cent. The children "completely cured of infection" showed an improvement of 60 per cent. The group of "severely infected children, treated but not cured" improved 38 per cent. An older subgroup of "severely infected children treated but not completely cured" improved 9 per cent less than the normal children

and much less than the untreated younger children. Dr. Strong concludes that "hookworm unmistakably affects mental development. Treatment alleviates this condition to some extent, but it does not immediately, at least, permit the child to gain as he would if he had not had the disease." His figures apparently show that "prolonged infection may produce prolonged effects on mentality—effects from which the individual may never recover."

CHAPTER I

THE PROBLEM, METHOD AND RESULTS. A GENERAL STATEMENT

The present study was undertaken with the desire to find out some of the factors that enter into temperament, moods or dispositions and to get some insight into any relation that may exist between mood and physical and mental performance. The subject is a difficult one because of the practically unexplored field of which it is a part, because of the involved nature of temperament itself and because of the few studies made of the subject by methods sufficiently objective in their nature to justify a fair measure of confidence in the results.

It was necessary at the outset to limit the field for investigation and to make a rather intensive study of some phases with the hope not so much of making any definite discovery as to point out what were well to do next and perhaps how *not* to do it. The tests employed were chosen from among those already used by psychologists to measure relatively small differences and which have been found to be susceptible to a wide variety of influences. These tests with slight modification, which will be discussed in detail later, were given to three different groups of individuals aggregating 18 women and 20 men, ranging in age from 20 to 51 with the median at 30 to 32. The tests for each group were given on the same day under uniform conditions for the group, at the same hours (between one o'clock and four o'clock p. m.), by the same person (the writer), and in the same order. Any departure from the order for the group, such as a slight difference in the time of testing any individual, was recorded but in no case did it extend beyond 45 minutes and is negligible.

Group I. The first series of tests began March 24 and ended April 28, 1920. The subjects were 3 women and one man, graduate students working at the time in the Columbia Psychological Laboratory, where this research was planned as part of the requirement of the course. The tests were made individually once a week at the laboratory period, covering a period of six weeks.

The subjects had a general idea of the purpose involved in the experiment and some of them had a preconceived idea that they were likely to prove less efficient when they were feeling below normal in cheerfulness.

Group II. The second series of tests were made on 10 subjects, five women and five men, patients in the Manhattan State Hospital, who had been under treatment for Manic-depression. At the time of the experiment these subjects were in a practically recovered state. Subjects of this type were chosen to take part in the experiment because of their susceptibility to changes in mood, thus insuring (to the experiment) a certain variety of records of mood.

The tests for the women began October 12 and ended November 18, 1920. Those for the men began January 12 and ended February 18, 1921. The tests were given on consecutive days four times a week (on Tuesday, Wednesday, Thursday and Friday), between one o'clock and four o'clock, under uniformly controlled conditions for each group, in the same order and by the same person (the writer). As in series I, any variation in the routine of testing did not extend beyond 45 minutes for any one subject and can be disregarded.

The records for the first week for some subjects were not entirely complete, owing to the necessity of trying out the individuals to determine their willingness and physical and mental fitness to take part in the experiment, and to establish a satisfactory routine. The selections were carefully made by members of the medical staff of the Manhattan State Hospital, who maintained throughout a generous professional oversight of conditions affecting the experiment.

In every instance the subjects were willing and glad to take part in the experiment. They were not aware of the object of it but they brought a peculiar sympathy to the work because of its evident relation to mental conditions. The spirit of all is expressed in the words of one of the subjects: "If it adds to knowledge about the mind, it will be a good thing to do and I'd like to help. It may be useful to other people with mental trouble."

In order to have the final figures represent the results of work done under conditions as nearly uniform as possible the scores secured for the first week were disregarded and those for the five weeks following were used as the basis of calculation.

Group III. The third series of tests began on February 28 and ended May 9, 1921. The subjects were 23 graduate students in psychology, 11 women and 12 men studying the problem of group testing in Columbia University. For this group obviously the tests given to Group I and Group II as individuals had to be modified. (See Chapter II for detailed description of test used and modifications made for Group III). The tests of pulse, blood pressure, strength of grip and accuracy of Touch Discrimination had to be omitted entirely. The Opposites Test was substituted for the Color Naming Test for the study of controlled association and the Free Association Test and the Tapping Test were modified also, to meet the conditions of a group test. These tests were given once a week at the same time of the afternoon, in the same order and by the same person (the writer).

The object of the experiment being to observe if mood was in any way related to performance, the method of rank difference was determined upon as possibly the best to use in the statistical treatment of the data secured in this experiment. Before undertaking the ranking of the scores both for mood and performance it became necessary to devise a method for ranking with a fair degree of certainty the objective marking for mood, and another for eliminating the effect of practice from the actual scores secured by the tests. The methods employed are described in detail in Chapter III. The scores for mood and for performance in each of the tests were correlated and ρ determined by the application of the formula,

$$\rho = 1 - \frac{6 \Sigma D^2}{N (N^2 - 1)}$$

By use of tables[34] values of r corresponding to values of ρ were found.

Summary and Conclusion.

I. The conspicuous feature of this investigation is the low coefficients that result from correlating measures of mood with tests of performance.

2. The results differ widely not only with regard to the performance of the separate individuals but with regard to the individual's performance in the several test.

3. The median coefficients for the three groups, men and

women combined, are practically zero for all of the fifteen tests.

4. The correspondence is fairly close between the actual distribution of the coefficients and the normal distribution for the sexes considered separately or combined in the ten tests that are more definitely tests of performance.

The median coefficient for both men and women is practically zero and the 25 percentile, the 75 percentile and Q are within .31 of zero for the men and women considered separately or combined. Fifty per cent of the coefficients lie between the distance + 1 Q and — 1 Q. The per cent of coefficients that lie within the remaining distances of Q is very close to the per cent of measures included within these distances in the normal curve of distribution.

5. No single test shows a sufficiently marked departure from the normal distribution to conclude that mood had influenced results.

6. There is little to indicate individual differences.

7. If a very marked relation existed between the mood and the performance of the individuals tested it is likely that it would have appeared to a more noticeable extent than it has in the results of this experiment. This statement of the situation is offered:

Extremes of mood act as distractions, the absence of any demonstrable effect on performance being due to the tendency of the individual to overcome resistance by increasing the output of energy.

This concluding statement is supported by the results of the well-known experiment of Morgan's on "Overcoming of Distraction and Other Resistances."[19] This experimenter found that the general tendency was for subjects to make movements of their speech organs as they performed tasks and, furthermore, there was a general tendency to increase these movements as they worked, as shown by the increase in the breathing rate. The subjects who made use of this means of overcoming the distracting effect of noises increased the quality of their performance and those who did not showed no such increase. In the test made on the typewriter it was found that the key pressure records showed that greater effort was exerted through the distraction period. The general tendency was to attempt to overcome distractions by strain and articulation and by these aids performance was kept

pretty well up to normal. In one part of the experiment Morgan observed the relation of time to force by having the subject raise varying weights with as much force as possible. The tendency was for the subjects to keep the time constant, extra force being exerted to meet the change in conditions.

It is common opinion that atmospheric conditions influence efficiency, yet the experiment by Thorndike on ventilation indicates that the individuals do as well in mental work under hot, humid, stale and stagnant air conditions as under the best conditions. The investigation made by Dr. Stecher show that varying degrees of humidity have practically no effect on efficiency. The school children tested by Thorndike after school hours agreed that they were tired, and thought that they could not work nearly so well, yet these same children *did* do just as well in the tests given. Dr. L. S. Hollingworth's investigation of the effect of functional periodicity did not reveal any mental or motor inefficiency in the women tested, contrary to general opinion that temporary discomfort or disinclination to action would tend to influence performance unfavorably.

Hookworm disease, bad tonsils and adenoids are drawbacks to performance, yet the investigations of Dr. Strong and Dr. Cobb do not reveal any marked increase in performance after the condition had been corrected. Equal performance under conditions of disease or temporary physical discomfort calls for increased effort on the part of the individuals. Certain investigators report marked improvement after the removal of tonsils and adenoids. The probability is that the operations resulted in improved health, which meant increased energy, less strain and more easily sustained attention to work.

Disease, temporary physical discomfort and mental fatigue are all conditions of disinclination to action which must be counteracted by voluntary effort if the individual's performance is kept up to normal. An unfavorable mood represents disinclination to action. When this disinclination is overcome by increase in the output of energy the influence on performance is slight.

The subject of mood furnishes a rich field for further experimentation. By way of practical suggestions the writer offers the following: There is a possibility that a truer ranking of mood might be obtained by using a measure that allows

for discrepancies between the objective marking and the subjects' description of their mood, especially where the mood is compared with that of the previous day. The comparison depends on memory and is likely to be somewhat inaccurate. It is possible, too, that more definite results might follow from the use of a limited number of tests covering in each case a longer time. By taking the record for mood at the beginning and end of the limited series of tests, changes of mood during the test would be indicated to some extent. It would be advisable at the outset to provide definite means for differentiating between depression and fatigue, depression and irritability or anger, and between cheerfulness and degrees of nervous excitability.

The experience of the writer in this investigation leads to the opinion that the relation between emotional disturbance and the accuracy of touch discrimination might be studied with profit; also the relation between emotional disturbance and the extent of voluntary movement[33] as measured by the drawing of a line to equal in length a remembered standard.

CHAPTER II

TESTS AND METHODS OF HANDLING DATA

SECTION 1

DESCRIPTION OF THE TESTS USED

The following tests were made in the order indicated below:

1. Measure of Mood.
2. Pulse Rate (Omitted in the group test).
3. Blood Pressure. (Omitted in the group test).
4. The Two Point Threshold. (Omitted in the group test).
5. Muscular Strength. (Omitted in the group test).
6. Line Drawn to Equal in Length a Remembered Standard.
7. Tapping Test.
8. Color Naming Test. (Opposites test substituted in the group test).
9. Addition Test.
10. Free Association Test.

1. Measure of Mood.

At the beginning of each experiment a piece of paper 8 in. x 10 in. was placed on the table in front of the subject. This paper contained a space at the top for the name and date and well down from the top of the paper a horizontal line 10 cm. long drawn in ink. This line occupied the middle of the paper, the margins being exactly equal. The subject was given the following directions: "If the left end of this line represents the lowest degree of cheerfulness you ever feel and the opposite end the highest degree of cheerfulness you ever feel, where on the line are you now in degree of cheerfulness? Draw a line across the scale to indicate where you are today." This line was drawn promptly, following a brief but noticeable introspection period.

The subject was asked to write a short description of his mood in the space below the scale. To this was added the observations and impressions of the operator regarding the physical or emotional condition of the subject at the time of the experiment. This step had to be omitted in the group test.

2. Pulse Rate.

In each case the count was taken at the wrist of the left hand, the subject being seated. For Group I the count was taken for one-half minute and doubled for scoring. For Group II the count was taken for one minute.

3. Blood Pressure.

This measure was taken for Group I and the women of Group II with the Baumanometer. For the men of Group II the Rogers Tycos was used. As the results of this study do not depend on comparisons between individuals but are concerned with the individual's range, any difference that might be due to a change of instrument can be disregarded. In every case the reading was taken with the subject seated and with the left forearm resting on the table, palm up. The stethoscope was placed just above the elbow on the brachial artery. The record for systolic pressure was in every case the *first* loud sound that was heard; for the diastolic the last faint sound that could be heard. In each instance there were two readings for verification.

4. The Two Point Threshold.

For determining the degree of sensitivity a simple aesthesiometer in use in Columbia College Laboratory was selected. It is made of a light weight Bristol board and constructed to admit of easy adjustment. The region tested was the back of the left hand about two inches from the knuckle of the third finger, the points of the aesthesiometer being placed just across the chord passing over the knuckle of the third finger, the strip of board connecting them being at right angles to the chord.

The subject sat opposite the experimenter with his left hand resting on the table, the hand being concealed from view by a screen constructed to stand across the lower forearm but nowhere to touch it. In every case the test began with a distance of two centimeters between the points of the aesthesiometer and this distance decreased gradually until a measure was reached when the responses of the subject began to be uncertain. The method of increasing the distance between the points stimulated was then applied and the smallest measure recorded when the responses were certain, as determined either by the method of increasing or decreasing the distance between the points stimulated.

5. Measure of Muscular Strength.

The arm dynamometer was used for Group I. It was attached to the wall and the subject stood at arm's length from the wall and pulled with his right arm. The use of this apparatus did not prove entirely satisfactory as it allowed for bodily weight to enter into the records, but as the factor entered into all records the results are usable for the group. For Group II the hand dynamometer was substituted for the arm dynamometer. The records were taken in every case with the subject standing and for both right and left hands.

6. Line Drawn to Equal in Length a Remembered Standard.

For securing a measure of the length of line that the subject would draw voluntarily the following method was used. At the time of the first experiment each subject was given a paper, 8 in. by 10 in. and was asked "to draw a line of convenient length anywhere on the paper." The paper was put aside and before the close of the experiment for the first day the subject was given another paper of the same size and grade as before and was asked to "draw a line as nearly equal in length as possible to the line you drew at the beginning of the test." At each test period he was asked to reproduce again the standard line he drew at the time of the first experiment.

7. Tapping Test.

A. For individuals.

The apparatus used for this experiment with Group I and II was the ordinary small tapping board in use in the Columbia Laboratory, connected up with an electric counter and a short metal rod for tapping. The subject sat before the board which was placed on a table, with the small metal rod held in his right hand and his right elbow resting on the table. The instruction was given, "Tap at a comfortable rate." At a signal from the operator he began tapping and continued tapping for one minute, time being kept by the operator with a stop watch, the tapping record being recorded by the electric counter. The time remained constant while the amount varied. It was observed that an occasional *tapping movement* owing to fatigue or lack of control due to other cause was not recorded because the movement fell *just* short of the brass plate. Any unrecorded tapping movement of this nature was kept account of and added to the score read from the electric counter.

B. For the Group.

Group III consisted of twenty-three individuals and this test had to be modified to meet the conditions of a group test and still have the results comparable with those of Group I and II where the tapping test was given as an individual test.

This test was given in the Columbia Psychological Laboratory with the subjects sitting on chairs at the laboratory tables. Before each subject was placed a piece of paper 8 in. by 10 in. on which were drawn ten horizontal lines about an inch apart. The subjects were provided with lead pencils of uniform length and size, selected to match as nearly as possible the metal rod used in the individual tapping test. The subjects were given this simple instruction: "This is to be a tapping test. The paper is your tapping board. The pencil is the tapping rod. Tap at a comfortable rate. As you tap will you tend to work along in either direction between the horizontal lines in order to distribute the dots, thus making the counting of them possible and easy. If this seems to hamper your movements disregard it." At a signal, as in the individual tapping test they began to tap and continued for one minute. The dots were well distributed along between the horizontal lines and taking them section by section, the count was easily made and easily summed. In all respects, in the opinion of the experimenter the test parallels the individual tapping tests, the only exception being the slight effort on the part of the subjects to distribute the dots. The subjects reported that the effort was so slight as to be entirely negligible.

8. Controlled Association.

A. Color Naming Test.

The Color Naming Test is designed to measure the speed with which the name or idea can be brought to consciousness and expressed vocally upon the sight of an object which in this case is a color. The subjects in Group I and Group II were provided with a Woodworth-Wells[32] Color Card, on which are printed in random order one hundred small squares in five different colors. The card was covered by a paper and placed on the table directly in front of the subject. At the first several trials each subject was asked to name the colors in the sample at the top of the card to make sure of his ability to discriminate the five colors. At a signal the operator re-

moved the cover from the card, passing it quickly to the subject's right and the subject began to name the colors as rapidly as possible from left to right. The time was taken with a stop watch to the fifth of a second. Errors, also, were taken account of. In the handling of errors the procedure varies from that generally in use where in the case of any misnaming of a color the subject is stopped and made to return to the point at which the mistake arose, all mistakes being accounted for in the time consumed by them. In the present experiment mistakes were kept account of by the operator, any omission being counted as a mistake. The object here was not only to discover if there was any variation from day to day but to discover in the case of color naming if it was in speed or accuracy. Furthermore, any interruption of the subject to point out his errors would tend to irritate and humiliate him thus introducing into the experiment avoidable complications. This variation seems not to be especially out of harmony with the spirit of the test as given by its authors: "The confusion here is sometimes due to wandering of the attention from the work in hand; but at other times it seems to be due to interferences generated by the performance itself. Whatever may be the cause of confusion in each particular case, efficiency in the test requires such a degree of control as will eliminate the confusion. Periods of confusion are but extreme manifestations of inefficient control; in a minor degree, the efficiency of one individual in comparison with another is shown by uniform slowness of response."[32] The confusion in this test due to whatever cause would be represented in the time taken for the completion of the test. The mistakes were kept account of and constitute a measure of accuracy.

B. Opposites Test.

The subjects of Group III were tested as a group. The Color Naming Test being an individual test, it was necessary to select an equivalent test of controlled association processes that could be administered as a group test. The Opposites Test of Woodworth and Wells[32] was selected. According to the authors this test is a "measure of mental alertness" and of "efficient mental control" and is "apparently the most available material for a test of completely controlled association, with the exception indeed, of the naming test and the arithmetic tests....."

Professor H. L. Hollingworth who has worked extensively with this test says: "This test indicates the ability of the individual to select the appropriate response from a host of ideas which follow in the wake of a stimulus word. It is an index of speed, accuracy, linguistic feeling and of ability to repress useless or irrelevant ideas. It is a test of association processes, but of association processes of a considerably more complex kind than those involved in the color naming."[12] This test then appears to be a fair substitute for the color naming test as a measure of controlled association. The fact that the association processes tested here are of a more complex kind than those brought into play in color naming need not be given much weight. The students who took part in this test were graduate students of psychology whose training doubtless gave them some advantage over the subjects in Group II who took part in the Color Naming Test. They were possibly at a slight disadvantage when compared with the subjects in Group I who also were graduate students in psychology. This group took the Color Naming Test.

The test of Hard Opposites was placed face down on the table in front of the subjects. At a signal the subjects turned the paper over and as rapidly as possible *wrote* the opposite to each word in order. Twenty seconds only were allowed for this test. Preliminary trials with the material indicated that thirty seconds were sufficient for some subjects to finish the test. Allowing for the time that would be gained by practice, twenty seconds semed a safe time to insure that none would exhaust the list before the time set for all. The two lists of Hard Opposites were given in alternation at each experiment period one week apart. The amount of work done and errors were recorded, the time being kept constant.

9. Addition Test.

The *constant increment* test devised by Woodworth and Wells[32] was the material used in making this test. The blank contains 100 two-place numbers arranged in four columns. The subjects of Groups I and II were provided with one column of the numbers. The slip was placed face down on the table before the subject and at the starting signal the subject turned the paper over and added 7 to each of the numbers recording the sum opposite each number in turn. In order to counteract as much as possible the effects of memory and practice, the remaining columns of figures were given in

order at successive test periods. The time taken to complete the column and the errors were recorded. Group III was provided at alternate test periods with half of the same test blank. One minute was allowed for performance. The amount of work done and the errors were recorded.

10. Free Association Tests.

Group I were given the first 25 words of the Kent-Rosanoff Association Test[17] for four successive test periods. The responses came almost automatically and were in every case common reactions. This was due in part to the fact that the subjects were familiar with the test.

Because of the rather wider range of the words included, the Woodworth Wells[32] list for the Free Association Test was used for the other two groups. The first 100 words were used, 25 being given at each test period the same 25 words being repeated at every fourth test period.

With Group I and Group II the stimulus word was pronounced by the operator the subject being required to "say the first word that came to mind other than the stimulus word." The response was recorded by the operator together with the time which was taken by the stop watch to one-fifth of a second. For Group III the procedure had to be varied slightly to meet the requirements of a group test. The subjects were provided with pencil and paper on which were numbered 25 spaces for responses. The stimulus word was given with a rising inflection by the operator and the subjects were required to "*write* the first word that came to mind other than the stimulus word." The stimulus words followed each other at exact intervals of five seconds by the stop watch. If no response came at the end of five seconds the space was left blank. At the end of the test any word not distinctly heard by any individual was repeated in the same manner as before and the response written by the subject, the subject being instructed to star all such words. In this test the time remained the same. The individual reactions are here the significant measure.

SECTION 2

Discussion of Methods of Handling Data. Scores for the Rating of Mood.

In order to preserve objectivity in the rating of mood, the scores were used as marked by the subjects. These scores were determined by measuring the distance in centimeters from the zero end of the scale to the point marked off by the subject as representing his degree of cheerfulness for each experiment period. These scores were placed in order beginning with the lowest and ending with the highest. In case of tying, and there was a fair number of tied scores, resort was made to the verbal description asked for after the subject had marked his mood on the scale. On the basis of these data the scores were ranked by the experimenter. It seemed desirable here, in order to counteract the personal equation in this ranking to secure the benefit of group judgment. The verbal descriptions of mood and the observations and impressions of the experimenter were copied on separate slips of papers, dated and submitted to five other judges, graduate students in psychology, who were asked to rank the tied scores. The judges did not know that the scores *were* tied. The averages of the six judgments were secured and ranked and the ranking was correlated by the method of rank differences with the ranking determined upon by the experimenter on the basis of what the subjects had said and the written record made at the time of the experiment.

In the case of one subject, F, Group II, when two scores were tied the judges' agreement with the writer's ranking was perfect. In another instance the same subject tied four times. There was no agreement and the scores were ranked as tied scores. In the case of Subject Km, Group II, there were three instances when two scores were tied. In all three instances there was perfect agreement between the writer's and the judges' rankings. In the case of Subject L, Group II, where four scores were tied there was perfect agreement. In a second instance of four tied scores by this same subject the agreement was .70. In four other instances when two

scores each were tied five out of the six judges gave the same ranking as the writer. Two scores for Subject S, Group II, were tied. There was perfect agreement among the judges regarding the ranking. The position of several other scores of this same subject were in doubt because of her failure to mark the scale for four days. The agreement here between the judges' arrangement and the writer's was .73. The results indicate a considerable correspondence in the opinion of the judges and also justify a measure of confidence in the method employed for measuring mood.

In two cases in Group II it seemed impossible to secure the marking of the scale to indicate the degree of cheerfulness. The idea of measuring a mental state by a space on a line seemed difficult for these subjects to grasp. Effort was made as the experiment progressed to secure the measure without antagonizing or irritating them but the effort proved unsuccessful to the end. The verbal descriptions given by the subjects of their mood and the experimenter's impressions of mood made at the time were available. With these taken as a basis for judging, the mood for the different days was ranked in order from lowest to highest. Again it seemed desirable in order to counteract the personal equation in the ranking to secure the benefit of a group judgment. The same procedure was followed as in the case of the tied scores, seven judges acting in these two cases. The average ranking of the seven judges when correlated with the ranking previously determined on by the experimenter gave in one case .78 and in the other .82. Subject B, Group II suffered a relapse into a manic state while the experiment was in progress. On recovering he invariably marked the scale at 100. On the basis of his statements and the writer's observations these scores were ranked by the writer. This arrangement correlated with the average rankings of six other judges gave a coefficient of .74. As in the instances of tying these fairly high correlations justify a measure of confidence in the original ranking, based upon the statements of the subject and the observations of the experimenter. While then the original ranking might be used with practically equal justification it seemed altogether desirable to use the judgment of the group as the basis of approximating the true ranking since the advantage of the group judgment over the individual judgment has been shown experimentally.

The question naturally arises, to what extent the reports of

depression and elation are clouded by the influence of anger and irritability at the one extreme or excitability at the other. There are in the verbal reports one instance of being "mad," one of being in a "bad humor" (the subject was evidently quite angry), one of being "cross," one "bored, with a calm acceptance of things." Among the observations made by the experimenter there is one instance of a subject being "irritable." The subjects in the group test when not cheerful reported "fatigue," "headache," "worry." There was one instance of "being cheerful despite headache." In the records for the upper level there was little to aid in differentiating excitability from actual cheerfulness. There were two instances of mixed emotion holding over from the day before when anger was felt because of an injury received. There had been a struggle between the spirit of anger and the virtue of charity and the latter was controlling at the time of the test. This is a small number out of the 413 records taken of mood. The report as a whole then may be considered as indicating levels of cheerfulness ranging from depression to mild elation, with some allowance made for records modified by anger or irritability.

SECTION 3

Method of Eliminating Practice Effect from the Original Scores

It is obvious that, in a repeated exercise of any of the functions brought into activity by the tests employed in this experiment, the effect of practice would be considerable. Despite the routine previously described which was employed in the Addition Test, the Opposites Test and the Free Association Test, namely of presenting in alternation tests of equal difficulty, thus tending to counteract the effect of practice, the scores showed the characteristic change that is conceded to be due to practice.

It is safe to assume that the effect of practice would not be present in the measures of Mood, Pulse, Blood Pressure, and Line Drawn to Equal in Length a Remembered Standard not seen from the time it was drawn.

For the rest of the tests, the Two Point Threshold Test, Tapping, Muscular Strength, Color Naming, Opposites, Adding and Free Association, the following method was employed for eliminating the effect of practice:

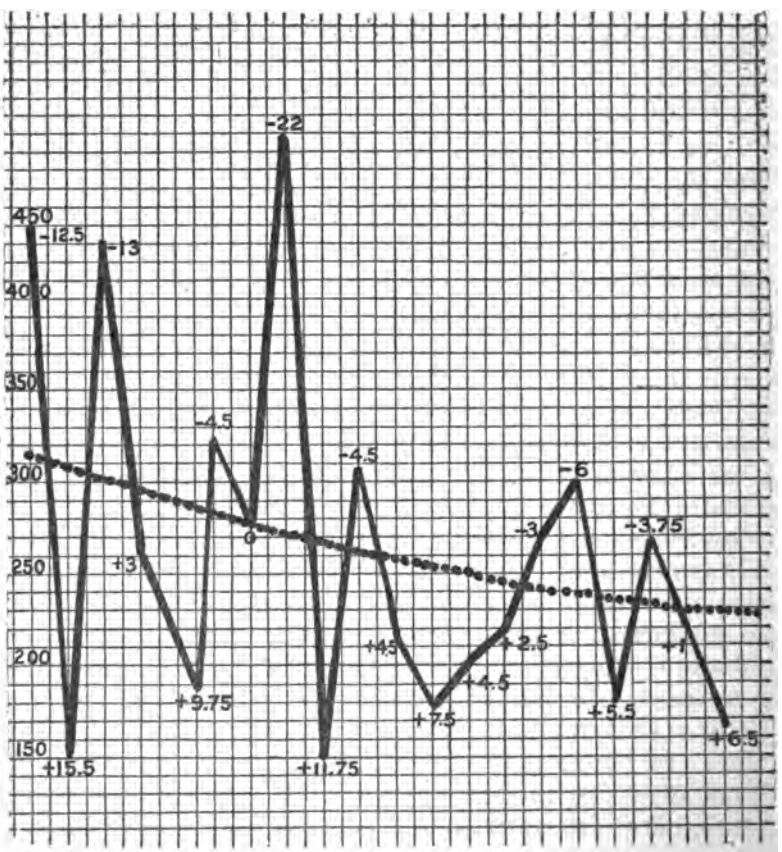
1. The curves for each of the series of test results were plotted on ordinary coordinate paper.
2. Investigation was made by the writer on the effect of practice for the various kinds of material employed. Fortunately for the present study results are available of thoroughgoing experiments covering not only *all* the material used but of experiments covering approximately the same length of time and in some cases a very much longer time. Many of the normal practice curves are to be found in Thorndike, "The Psychology of Learning," Vol. II., Chap. VI and VII. These were copied out and additional curves plotted from the results of Woodworth and Wells, of Poffenberger[21] and of Hollingworth[13] because of the very close correspondence of the present investigation as regards material, method employed and time covered.
3. Using these practice curves as a criterion, the writer drew a practice curve for each series of test results previously plotted. The resulting curve in each case presented the

characteristic features of the normal practice curve, modified to conform to the *individual's* performance in each *individual* series of tests.

4. In order to offset the personal equation of the single judge the judgment of two other people was secured for each curve. One Doctor of Philosophy and two graduate students, assistants in the Psychology Department at Columbia, acted as judges. In order to keep the influence of suggestion out of the results the writer drew in the first curve on a sheet of tissue paper which exactly fitted the coordinate paper on which the test results were plotted. This paper was carefully labeled and put aside. The original curves with a fresh tissue carefully labeled and with definite instructions as to placing it, were given to each judge. He was provided also with the copy of the principal contributions to the literature concerning practice curves (see 2 above) which he was instructed to review before representing his judgment of the shape of the curves for the material under consideration. The tissue record of this judge was put aside and the same procedure was carried out with the next judge.
5. The three sheets of tissue were then placed together on the original curves, and a line representing the average curve for the three judges was drawn in. Preponderance of weight was given to the judgment of two as against one, the position of the average curve being raised or lowered according to the distance two judges stood above or below the third. The position of this average curve thus carefully determined throughout its entire length, it was marked in by a broken line to keep it free from the possibility of being confused in any part of its course with the continuous lines used by the judges in representing their judgments. By means of a carbon paper placed face down under the lower tissue this average curve was readily transferred to its place on the original curve. This was in all cases the only line appearing in connection with the original curve and represented the average opinion of three judges as to the position of the normal practice curve for the individual in the particular test in question. In other words, it represents the position of the scores the individual in question could have been expected to have made, unmodified by other factors. Any deviation from this normal practice curve,

modified to suit the individual performance in any particular test, can reasonably be said to be due to the influence of some other factor and not to practice effect. It is clearly seen that any correlation to be reliable must be made between mood and scores which represent the degree of deviation from the normal practice curve.

6. Accordingly, for each score a derived score was obtained by counting on the original plotting the number of squares above or below the point in the normal practice curve that any individual score lay. See curve that follows.



Curve for Addition
Subject P. Group II.
1 sq. = 10 Seconds.

Score 1 is 12.5 squares above the normal practice curve, hence is marked .–12.5. Score 2 is 15.5 squares below, hence is marked +15.5. The score for the fifth day is 9.75 squares below, hence is marked 9.75. Score 7 is at the point where it could be expected to be on the seventh day of practice, hence is marked 0. The score the eighth day is 22 squares above, hence is marked –22 and so on. These derived scores, determined in terms of squares, represent the individual's performance, with practice effect eliminated. They were then substituted in every case for the original scores and used as the basis of all calculations of tests where practice effect obviously entered.

The fact must not be overlooked that variability decreases with practice and that deviations of a given amount become more significant as practice proceeds. The derived scores used as a basis of calculation must be considered in relation to the ratio of variability that exists between deviations in the earler and later days of practice. This ratio may be determined by dividing each deviation by the distance that the point on the type curve from which it is calculated is from zero. In the earlier days of practice (in a practice curve for *time* which is used here for illustration) the distance from the type curve to zero is larger than in the latter days of practice. This distance for each day becomes the denominator, the deviation for each day the numerator from which the ratio may be calculated. In the earlier days of practice the denominator is relatively larger, hence the ratio derived by dividing is smaller. In the later days of practice the denominator is smaller and while the deviations normally are relatively smaller the ratio becomes relatively larger. By calculating the ratio between the distance from zero of any point on the type curve and the deviation from that point the relative value of any deviation may be determined.

In the accompanying curve for the first day of practice the distance from zero of the first position on the type curve is 315; the deviation for that day is –12.5, which gives a ratio of .0397. For the second day the distance of the type curve from zero is 305; the deviation is + 15.5, which gives a ratio of .0508. For the third day the distance of the type curve from zero is 300; the deviation is –13 which gives a ratio of .0433. The ratio for each day of practice being calculated the ratios are ranked from the largest minus to the largest plus and

this order should be used to determine the degree of correlation with the rankings for mood.

In this study the absolute and not the relative value of deviations were used in calculating the coefficients of correlation. In order to test the validity of these coefficients the ratio of the deviations were worked out for a number of curves. Care was taken to select cases that might be affected by using relative deviations. These selections included curves which indicated that the effect of practice was considerable; others where the effect was slight and still others where the scores made in the later days of practice were actually inferior to those made in the earlier part of the practice. The ratio of variability for each day of practice was calculated from these curves. The ratios were ranked and correlated with the rankings of the absolute deviations used in this study and in no case did the coefficient of this correlation fall below .98. The coefficients of correlation that appear in Tables I, II, III, XVII, XVIII, XIX of this study are then practically what they would be if calculated from relative deviations.

The following method was used, also, to determine the ratio of deviation at different points on the practice curve. The average deviations were secured for successive groups of five. In order to keep the change gradual the average was found for the first five deviations; then for the second to the sixth inclusive; next for the third to the seventh inclusive, and so on, to the end of the curve. The ratio was then found between successive average deviations and the actual deviations. These ratios were ranked in order from the largest minus to the largest plus. By this method due weight could be given to fluctuations in the later periods of practice. The method is applicable to curves plotted for amount of work done and to curves plotted for time taken to do work. The rankings secured by this method correlated practically perfectly with the rankings of the absolute deviations.

CHAPTER III

DISCUSSION OF RESULTS

SECTION 1

THE RELATION OF MOOD TO PERFORMANCE IN THE VARIOUS TESTS

The object of this research being to observe if degrees in cheerfulness parallel quantity and quality of performance the method of rank difference was determined upon as possibly the best to use, considering the nature and purpose of the experiment. The scores for mood were ranked according to the method described in Chapter III. The derived scores, determined by the method described in the same chapter, were ranked in order from the lowest minus to the highest plus. The relationship between the two rankings was then determined in terms of relative position by the method of rank differences employing the formula,

$$\rho = 1 - \frac{6\Sigma D^2}{n(N^2-1)}$$

The correlation coefficients thus secured, based on rank difference, are almost necessarily used when the variables to be correlated are expressed in ranks. For purposes of ready comparison r, corresponding to ρ, was obtained by use of tables devised for this purpose. The maximum change from ρ to r is .018, a difference that may be disregarded in the present study which is not concerned with fine distinctions but with tendencies.

It should be explained that the test results in every case were arranged so that a positive correlation always means the correlation between good performance and cheerful mood. In measures of speed and amount the scores were arranged from lowest to highest. In the case of accuracy of course the best records were those where fewest errors were made. In the test for Circulation the scores were arranged from lowest to highest. In results for Line Drawn to Equal in Length a Remembered Standard longer lines always mean good performances. Possibly if we knew what was being measured the drawing of shorter lines might be the desirable perform-

ance, in which case the sign of the correlation should be changed. The same is true for the tests of Circulation. This possibility must be borne in mind in interpreting the data for the tests.

The following tables give the 308 coefficients of correlation of mood with performance for the three groups in each of the tests:

TABLE I. GROUP I

	Pulse Rate	Systolic Blood Pressure	Line	Two Point Threshold	Right Hand Grip	Tapping	Color Naming	Errors	Addition	Errors	Individual Reactions	Average Reaction Time
					Women							
L.	.00	-1.00	.21	-.42		.00	-.66	.42	-.82	-.42	.52	.14
K.	.05	-.20	.21	.72	.00	.72	-93	.72	.56	.92	.92	.96
M.	.15	.35	-.62	-.15	.67	.38	.03	-.46	-.64	-.23	.05	-.10
					Men							
C.	-.52	-.52	.52	.15	1.00	.50	.52	-.64	.52	.53	.52	-1.00

TABLE III. GROUP III

	Line	Tapping	Opposites	Errors.	Additions	Errors	Individual Reactions
			Women				
H.	.26	.10	.20	.52	-.15	.52	-.03
L.	-.56	-.48	-.34	.52	-.22	-.05	-.40
McN.	-.28	-.50	.26	-.15	.02	.81	-.10
M.	-.05	.05	.02	.52	-.25	-.17	.06
Ti.	.21	-.36	-.68	.52	-.17	-.01	-.71
Co.	-.15	-.21	-.96	.52	.27	-.28	.35
R.	.90	-.50	-.31	.50	.40	.50	-.74
Ca.	-.33	-.44	.34	.21	.37	-.62	.13
D.	.16	-.81	-.81	.50	-.67	.80	-.57
Ta.	.62	.21	.81	.50	-.16	.00	-.86
O.	-.12	-.21	-.04	.52	.18	.46	.35
			Men				
Ac.	-.23	-.84	.03	.52	.16	.74	.14
Al.	.77	.82	.04	.60	.60	-.12	-.44
Ax.	-.11	.00	.17	.50	-.06	.23	.31
Ga.	.91	.62	-.59	.52	-.62	-.10	-.15
Ka.	.19	-.72	-.83	.52	.08	.10	-.33
Ko.	.52	-.26	.28	.52	-.26	-.25	-.52
Ma.	-.05	.42	.16	.52	.40	.19	.66
Mi.	-.44	.67	.66	-.15	.05	.52	.40
W.	-.38	.82	.20	.52	.91	-.46	-.68
J.	-.35	-.36	.21	.52	.16	-.15	-.44
Gr.	.48	-.88	-.86	.52	.70	.58	-.32
N.	-.31	-.31	.57	.52	.40	-.36	.60

TABLE II—GROUP II

	Pulse Rate	Systolic Blood Pressure	Diastolic Blood Pressure	Pulse Pressure	Line	Two Point Threshold	Right Hand Grip	Left Hand Grip	Tapping	Color Naming	Errors	Addition	Errors	Individual Reactions	Average Reaction Time
Women															
L.	.29	-.27	-.33	.02	-.31	.27	.52	.66	.25	.21	-.27	-.29	.29	.29	.25
S.	-.49	.44	.13	.01	-.72		-.37	-.33	.15	-.08	-.15	-.08	.21	-.29	-.34
K.	-.10	.06	-.15	.08	.33	.23	-.13	-.13	-.08	-.15	.13	.54	.15	-.31	-.31
H.	-.02	-.19	-.02	-.17	.24	.46	.13	-.13	.42	.23	.02	-.17	-.42	.20	.29
P.	-.10	.33	.37	-.08	.21	.02	.33	.54	.74	.35	.02	.23	.08	.15	.31
Men															
F.	.17	-.08	.05	.10	.17	.17	-.10	-.70	-.10	-.06	-.37	-.46	.79	-.02	-.17
B.	-.08	-.47	-.61	.48	-.40	-.37	.37	.46	-.51	-.50	.10	-.46	.19	-.33	-.37
O.	-.04	.02	-.04	-.15	-.25	-.15	-.10	-.25	.08	.27	.04	-.08	.35	.31	-.17
T.	-.10	-.04	-.10	.08	-.04		-.33	-.06	.21	-.08	-.13	.27	.01	-.21	-.04
K.	-.02	-.50	.29	-.35	.21	.02	.23	.64	.44	.13	-.35	-.33	-.10	-.08	-.25

Because of the confidence with which the median is accepted as an index of tendencies, it is selected as the basis for further consideration of agreement in results. The median coefficients for the three groups of subjects are given in the table that follows, together with the reliability of each figure, computed by means of the formulae,[27]

$$\sigma\text{dis.} = \sqrt{\frac{\Sigma d^2}{n.}}$$

$$\text{P. E.}_{\text{t. med.} - \text{obt.med.}} = .6745\left(\frac{5}{4}\frac{\sigma\text{dis.}}{\sqrt{n.}}\right)$$

In all instances the figures for reliability mean that if the experiment were repeated in the same way and under the same conditions what are the chances out of 100 that the median will be equal to or greater than zero. It will be noticed that the reliability is low in a few cases, fair in others, and high in still others, but in *all* cases with one exception the probability is that the obtained median coefficient approximates toward the true median coefficient, especially with respect to its position away from zero.

On examining the results presented in the following tables it is readily seen that they differ widely not only with regard to the performance of the separate individuals but with regard to the individual's own performance in the several tests. For example, for Group I, the coefficients of correlation for mood with pulse vary from .00 to .15 with the median at .05 for the woman. If the results for the one man who took part in this division of the experiment were included, the range would extend from –.50 to .15 with the median at –.02. In Group II the coefficients of correlation for the women range from –.47 to .28 with the median at –.10. For the men of this group the range for pulse is from –.08 to .17 with the median at .04.

For the Two Point Threshold Test, the range for Group I is from –.42 to .72 with the median at –.15. The individual coefficient of correlation for the man of Group I is .13, slightly above the median for the women of this group. The women of Group II range from .02 to .46 with the median at .25; for the men of this group the range extends from –.37 to 17 with

Median Coefficients

	Pulse Rate	Systolic Blood Pressure	Diastolic Blood Pressure	Pulse Pressure	Line	Two Point Threshold	Right Hand Grip	Left Hand Grip	Tapping	Color Naming	Errors	Addition	Errors	Individual Reactions	Average Reaction Time
						Women									
Group I	.05	.20	.02	.01	.21	.15	.33	.13	.00	.03	.42	.64	.23	.52	.14
Group II	.10	.06	.02	.01	.21	.25	.13	.13	.25	.21	.02	.08	.15	.15	.25
Group III	.01	.05	.02	.01	.05	.23	.13	.13	.36	.04	.50	.15	.00	.10	.195
Combined Average	.02	.06	.00	.03	.21	.18	.31	.12	.08	.03	.50	.15	.08	.05	.14
S. D.	.21	.03	.44	.087	.40	.23	.35	.39	.07	.48	.27	.05	.13	.45	.39
P. E. of Med.	.06	.0084	.1686	.033	.0786	.0745	1.2	.92	.41	.21	.33	.38	.42	.58	1.5
Med./P. E.	.16	5.5	.11	.33	2.7	3.1	.1058	.1494	.99	.0941	7.00	2.1	.97	.0859	.1263
Reliability	47.33	100.	47.31	58.02	96.57	58.17	79.58	27.09	27.09	55.36	.0659 100.	.0745 7.83	.0821 72.91	63.21	84.41
						Men									
Group I	.04	.08	.04	.08	.04	.065	.10	.06	.08	.06	.13	.33	.19	.08	.17
Group II	.04	.06	.04	.01	.08	.05	.10	.13	.13	.165	.52	.16	.00	.215	.17
Group III	.14	.08	.14	.08	.04	.065	.10	.06	.00	.13	.52	.08	.10	.15	.17
Combined Average	.01	.21	.08	.05	.04	.08	.02	.02	.03	.01	.28	.09	.12	.06	.20
S. D.	.30	.22	.29	.26	.20	.20	.25	.48	.53	.45	.33	.41	.35	.39	.10
P. E. of Med.	.0145	.0843	.1116	.0996	.0824	.0843	.067	.1839	.1069	.0923	.0678	.0842	.0719	.0802	.0383
Med./P. E.	2.7	.94	.71	.82	.60	.71	1.5	.31	.28	1.4	7.6	.94	1.3	1.8	4.4
Reliability	3.42	27.09	31.84	70.53	34.21	31.84	15.59	41.89	50.00	82.75	100.00	72.91	80.97	12.04	00.15
						Subject only.									
Men and Women															
Combined Average	.02	.08	.03	.015	.045	.02	.065	.095	.05	.035	.50	.06	.09	.055	.10
S. D.	.01	.12	.04	.01	.02	.07	.10	.13	.05	.02	.28	.02	.13	.06	.01
P. E. of Med.	.157	.37	.28	.19	.39	.32	.27	.37	.60	.47	.22	.40	.41	.42	.36
Med./P. E.	.0301	.0866	.0763	.0516	.0565	.0817	.066	.1006	.0843	.066	.0309	.0593	.0617	.095	.0843
Reliability	.63	.92	.53	.20	.83	.24	1.5	1.06	.59	.53	13.00	1.01	1.5	.93	1.1
	34.29	27.09	36.79	58.02	70.53	55.36	84.41	25.00	36.79	63.21	100.00	75.00	84.41	27.09	32.91

the median at –.065. In the cases of one man and one woman of the group it was impossible at any time to obtain a reliable record of ability to discriminate the two points of the aesthesiometer. Their responses to stimulation were strangely wide of the truth and appear to be unaffected by any increase or decrease in distance between the two points simultaneously stimulated. Stimulation of points varying in distance from 2 cm. to 7cm. apart on the back of the hand, palm of the hand, forearm, tips of fingers, lips or forehead failed to result in any more accurate responses. The responses varied from 3 cm., 4 cm., to 38 cm. for simultaneous stimulation of two points. Indeed, the same responses seemed to follow with equal variety if only one point was stimulated. The inability to discriminate the two points seemed to bear no relation to mood or ability to react with varying degrees of accuracy to the other tests given from day to day.

The results for tapping show a range in coefficients for the women of Group I from –.38 to .72 with the median at .00. The correlation coefficient for the men of this group is .50. The women of Group II vary from –.08 to .74 with the median at .25. The men of Group II vary from –.57 to .44 with the median at .08. The coefficients for the women of Group III range from –.81 to .05 with the median at –.36, for the men of Group III from –.84 to .00 with the median at –.13. Correlation coefficients indicating as wide a range appear in the results for the Addition Test, for the Color Naming Test; for Free Association, for Tests of Muscular Strength and the Line Drawn to Equal in Length a Remembered Standard.

A few high correlations appear in the results. For example, in the case of K., Group I, correlation of mood with tapping and color naming give coefficients of .70 and .90 respectively. Subject S., in Group III, shows a coefficient of –.70 and .90 respectively. Subject S., in Group III, shows a coefficient of –.72 for the correlation of mood with Line Drawn to Equal in Length a Remembered Standard. The coefficient for Subject K., Group II, in the Addition Test, is .54; for Subject P. in the same group, the coefficient for the tapping test is .74.

In the results for Strength of Grip the comparatively high coefficient of –.70 in the case of Subject F. and .64 in the case of Subject K. for left hand grip are in interesting contrast to -.10 and .22, respectively, for the right hand grip. Noticeable in the results for Group III is the coefficient of .82 in the Tap-

ping Test in the case of Subject W; of –.72 in the same test for Subject Ka; of –.88 for Subject G. Other instances are easily apparent on inspection.

Any single high correlation does not mean a great deal considering the period of time covered by the experiment. It is obviously a difficult matter to find people who can spare time regularly and for any extended period to take part in an experiment of this nature. This experiment with Groups I and III had the advantage of having subjects who worked at the regular recitation periods. It had the disadvantage of being interrupted in each case by the spring vacation and of being terminated automatically by the closing of the college term. With the subjects of Group II it seemed advisable to limit the testing to six weeks in order not to impose unduly upon the courtesy of the hospital, and on the time of the medical staff, nurses and attendants, and above all, on the patience of the subjects themselves.

The Probable Error of the various coefficients presented in the above tables is large, yet the results are not entirely without significance when it is remembered that they represent calculations from the scores made by a variety of subjects, both men and women, under controlled conditions as to place of giving the tests, method adhered to, and material used.

The method used for measuring mood has certain obvious weaknesses but in view of the test made of it by the method detailed in Chapter II, in the opinion of the writer it can be accepted as a rough measure of mood, especially in a study that is primarily concerned with relative positions indicative of tendencies and not with fine distinctions.

Considering the test results as a whole, for the women, out of 160 correlation coefficients between mood and performance, 88 or 55 per cent are positive, 70 or 44 per cent are negative, while 2 or 1 per cent are zero. For the men, out of 148 correlation coefficients, 79 or 53 per cent are positive and 69 or 47 per cent are negative. For the group as a whole, out of 308 correlation coefficients, 167 or 54 per cent are positive, 139 or 45 per cent are negative, and 2 or less than 1 per cent are zero.

The last ten tests are more definitely tests of performance than are the first five tests (for Circulation and Drawing of a Line to Equal in Length a Remembered Standard). The 248 coefficients of correlation for these ten tests were plotted in the form of a distribution curve. Table IV gives the results

of this distribution—the 25 percentile, the 75 percentile, the median, the P. E. of the median, Q and the range. Table V shows the correspondence of the actual distribution to the normal distribution. The correspondence is fairly close between the actual and the normal distribution for the sexes considered separately or combined. There is no clear departure from zero correlation which is here assumed to be the norm and consequently little to indicate the influence of mood.

TABLE IV

	25 Percentile	Median	P. E. of Median	75 Percentile	Q	Range	No. of Coefficients
Women	−.28	+.046	±.03	+.268	.274	−.96 to +.96	130
Men	−.312	+.022	±.036	+.316	.314	−100 to +1.00	118
Men and Women Combined	−.3	+.036	±.0026	+.3	.3	−1.00 to +1.00	248

TABLE V

	−4Q	−3Q	−2Q	−1Q	+1Q	+2Q	+3Q	+4Q
Normal Distribution	.012	.067	.16	.50	.16	.067	.012	
Women		.065	.187	.50	.163	.083		
Men	.005	.072	.172	.50	.157	.088		
Men and Women Combined	.0081	.081	.1613	.50	.137	.081	.024	

TABLE VI

	25 Percentile	Median	P. E. of Median	75 Percentile	Q	Range	No. of Coefficients
Tapping	−.395	+.0375	±.08	+.387	.389	−.88 to +.82	37
Controlled Association	−.337	+.087	±.068	+.371	.354	−1.00 to +.92	37
Addition	−.243	+.16	±.063	+.374	.308	−.82 to +.91	37
Errors in Addition	−.168	+.112	±.07	+.514	.341	−.67 to +.92	37
Individual Reactions	−.393	—.05	±.07	+.318	.3555	−.74 to +.92	37

TABLE VII

	−4Q	−3Q	−2Q	−1Q	+1Q	+2Q	+3Q
Normal Distribution	.012	.067	.16	.50	.16	.067	.012
Tapping		.081	.162	.513	.216	.027	
Controlled Association		.108	.135	.648	.081	.027	
Addition		.108	.135	.518	.216	.027	
Errors in Addition		.027	.243	.486	.216	.027	
Individual Reactions		.027	.216	.54	.189	.027	

Table VI shows the results of a similar distribution of the 37 coefficients of correlation for each of the five tests in which all of the 37 subjects took part. Table VII shows the correspondence between the actual and the normal distribution. Table VI shows the range to be practically the same for all tests. The 25 percentile and the 75 percentile are also about the same for Tapping, Controlled Association and Individual Reactions. In the Addition Test, both for time and errors, the measures tend to fall toward the positive side of zero but still not far from zero.

Compared with the normal distribution, the actual distribution shows that the measures are slightly more numerous toward the center in the Controlled Association Test and in the test for Individual Reactions, and tend also to be in excess of the normal distribution between +2Q and +1Q in Tapping, Addition, Errors in Addition and Individual Reactions. The distance, —1Q and —2Q shows an excess of measures for Errors in Addition and Individual Reactions. In no case, however, is the departure from the normal distribution sufficiently marked to conclude that mood had influenced results. Taking the individual separately in these ten tests of performance, there are but two whose r is always either plus or minus and 12 who are consistently plus or minus from 80 to 100 per cent of the time; so there is little to indicate individual differences.

The question arises to what extent the low correlation coefficients are due to inferior performance in the lower and upper levels of cheerfulness, the middle levels of cheerfulness, perhaps, being favorable to better performance. If such were the case the coefficients of correlation would be low yet there would actually be correspondence between mood and performance. To test this the rankings for mood and for each of the tests for all the subjects were divided into four levels and the median rank determined for each level. There were a

few instances of direct positive relationship between levels of mood and levels of performance as observed from the median rank for each level. A few direct negative relationships were also observed. In every instance this same relationship was shown in either a positive or a negative coefficient as a result of correlating the actual rankings. There was no individual whose performance was consistently best or worst at any one level. The middle levels of mood did not show that performance at these levels tended to be superior to that in the higher or lower levels of mood. In fact the relations observable in this arrangement of the data would indicate a low degree of correlation between mood and performance.

CHAPTER III

SECTION 2

Combination of Tests to Secure a Measure Freer from Attenuation

The original purpose in working out the correlations was to analyze out some of the factors which enter into mood and to do this a rather wide variety of tests were used and the results correlated with the mood taken at the time of performance. It seems quite possible in correlating the results of the tests separately with mood which at best must be rather a rough measure, that any relation present would tend to be minimized and result in a low coefficient.

Accordingly the results of several tests obviously measuring common elements were combined in an effort to get a measure freer from attenuation due to chance errors which do not balance out in the case of correlation but cause the coefficient to approximate zero. The following tests were combined because of the obvious common element:

1. For Muscular Strength.....Right and Left Hand Grip
2. For Speed—Color Naming (time) and Tapping.
3. For Speed—Color Naming, Adding (time) and Average Reaction Time in the Free Association Test.
4. For Accuracy—Color Naming, Adding and Individual Reaction, errors taken in all three cases.

In each of the four combinations the rank orders as determined for the derived scores were summed. The composite measure thus obtained was ranked and as was the case with the individual tests correlated with the rankings for mood by the Method of Rank Difference. Tables IX and X show the corrected coefficients and also the coefficient for the individual tests combined to secure the composite measure.

Three tendencies are apparent in the corrected coefficients:

1. A tendency for the positive coefficients to increase in the positive direction.

2. A tendency for the negative coefficients to increase in the negative direction.

TABLE IX

CORRELATION COEFFICIENTS FOR COMPOSITE MEASURES

		Muscular Strength Right and Left Hand Grip		Speed Color Naming and Tapping		Speed Color Naming Addition Average Reac. Time		Accuracy Color Naming Addition Individ. Reactions	
					.70		.90		.70
Group I Women	K.			.97		.75		.80	
					.90		.55		.90
	M.			−.14	−.37	−.41	.03	.00	−.45
					.03		−.63		−.22
	L.			−.50	.00	−.80	−.65	−.20	.40
					−.65		−.80		−.40
							.21		−.27
Group II Women	L.	.68	.50	.15	.21	−.04	−.29	−.23	.21
			.64		.25		.25		.29
	S.	−.54	−.37	.06	−.08	−.21	−.08	−.25	−.15
			−.33		.15		−.08		.21
							−.34		−.29
			−.12		−.14		−.14		.12
	K.	−.14	−.12	−.20	−.08	.08	.52	−.04	.14
							−.30		−.30
			.12		.22		.22		.02
	H.	.00	−.12	.44	.40	.15	−.16	−.17	−.42
							.28		.20
			.33		.35		.35		.02
	P.	.56	.34	.52	.74	.40	.23	.28	.08
							.31		.15
			−.10		−.06		−.06		−.36
Group II Men	F.	−.54	−.68	−.20	−.10	−.34	−.44	.16	.78
							−.17		−.02
			.36		−.48		−.48		.10
	B.	.42	.44	−.64	−.49	−.66	−.44	.18	.18
							−.37		−.33
			−.10		.27		.27		.04
	O.	−.19	−.24	.24	.08	−.19	−.08	.29	.35
							−.17		.31
			−.33		−.08		−.08		−.13
	T.	−.31	−.06	.06	.21	.00	.27	−.06	.01
							−.04		−.21
			.23		.13		−.13		−.35
	Km.	.44	.64	.42	.44	−.34	−.33	−.42	−.10
							−.25		−.08

TABLE X

CORRELATION COEFFICIENTS FOR COMPOSITE MEASURES

Group III	Speed Tapping Opposites Addition	Accuracy Opposites Addition Individual Reactions	Group III	Speed Tapping Opposites Addition	Accuracy Opposites Addition Individual Reactions
	Women			Men	
		.10 .50	Al.	.48 .04	-.40 .11
H.	.28	.19 -.03 .50		.60	-.44
		-.14 -.03		-.82	.52
		-.46 .50	Ac.	-.30 .08	.43 .74
L.	-.75	-.38 -.33 -.05		.16	.14
		-.21 -.38		.00	.50
		-.48 -.14	Ax.	.07 .17	.38 .23
Mc.	-.21	.25 .00 .80		-.06	.31
		.02 -.10		.60	.50
		.03 -.17	Ga.	-.26 -.57	-.30 -.19
M.	-.08	.02 -.15 -.17		-.60	-.14
		-.24 .06		-.87	.68
		-.20 .44	Gr.	-.59 -.85	-.04 .56
O.	-.25	-.04 -.37 .44		.68	-.31
		.17 .34		-.35	.50
		-.35 .50	J.	.30 .20	-.57 -.14
Ti.	-.69	-.66 -.61 -.01		.15	-.42
		.16 -.69		-.70	.50
		.20 .50	Ku.	-.59 -.82	-.10 .10
Co.	-.89	-.95 .44 -.27		.08	-.32
		.26 -.34		-.26	.50
		.48 .50	Ko.	-.04 .28	-.71 -.24
R.	-.35	-.30 -.52 .50		-.26	-.50
		.38 .44 -.72		.67	-.14
		-.42 .20	Mi.	.73 .66	.15 .50
Ca.	.33	.33 -.49 -.62		.05	.38
		.36 .12		-.31	.52
		-.80 .50	Ni.	.19 .57	.08 -.36
D.	-.67	-.80 -.57 -.65		.40	.60
		-.65 -.65		.40	.50
		.20 .50	Ma.	.77 .15	.69 .18
Ta.	.21	.80 .36 .00		.38	.64
		-.15 -.85		.82	.50
			W.	.54 .20	-.64 -.44
				.91	-.66

3. A tendency for both positive and negative coefficients to take a position somewhat between the raw coefficients or even beyond either extreme, whether the raw coefficients were all positive, all negative or both positive and negative.

To bring together this loosely connected group of correlations in any sort of consistent scheme is almost impossible but it can certainly be said that taking account of attenuation by the method used here has in a number of cases increased the correlation coefficient by extending it in either a positive or negative direction. In other cases it resulted in bringing the coefficient to somewhat of an intermediate position. In a

TABLES XI

EXAMPLES OF THE FIRST TENDENCY

Tests Combined	Subject	Group	Raw	Coefficients		Corrected Coefficients
Rt. Lt. Hand	L	II	.50		.64	.68
Rt. Lt. Hand	P	II	.33		.54	.54
Color Naming and Tapping	H	II	.22		.40	.44
Color Naming and Tapping	K	I	.70		.90	.97
Color Naming, Adding & Ind. R.	P	II	.35	.23	.31	.40
Tapping, Opposites and Adding	Mi	III	.67	.66	.05	.73
Tapping, Opposites and Adding	Ma	III	.40	.15	.38	.77

TABLE XII

EXAMPLES OF THE SECOND TENDENCY

Tests Combined	Subject	Group	Raw	Coefficients		Correct Coefficients
Rt. & Lt. Hand Grip	S	II	−.37	−.33		−.54
Color Naming and Tapping	F	II	−.06	−.10		−.20
Color Naming and Tapping	B	II	−.48	−.49		−.64
Color Naming, Adding, & Ind. Reactions (Errors)	Km	II	−.35	−.10	−.10	−.42
Opposites, Adding & Indiv. Reaction	L	III	−.40	−.33	−.21	−.75
Color Naming Adding Avg. Reaction Time	B	II	−.48	−.44	−.37	−.66
Tapping, Adding Indiv. Reaction	T	III	−.35	−.66	−.16	−.69

TABLE XIII

EXAMPLES OF THE THIRD TENDENCY

Tests Combined	Subject	Group	Raw	Coefficients	Corrected Coefficients
Color Naming and Tapping	L.	II	.21	.25	.15
Color Naming and Tapping	O.	II	.27	.08	.24
Tap. Add'g and Opposites	Al.	III	.32	.04 .60	.48
Tapping and Opposites (Time)	W.	III	.82	.20 .91	.54
Opposites, Add'n, Ind. Reactions	Ak.	III	.52	.74 .14	.43
Rt. & Lt. Hand Grip	F	II	–.10	–.68	–.54
Tapping, Opposites, Adding	McT	III	–.30	–.30 –.65	–.67
Color Naming and Adding, Avg. Reaction Time	F	II	–.06	–.44 –.17	–.34
Opposites, Adding, Indiv. React.	J.	III	.50	–.14 –.42	–.57
Opposites, Adding, Indiv. React.	G.	III	.50	–.14 –.10	–.30
Opposites, Adding, Indiv. React.	J.	III	.50	–.14 –.42	–.57
Tapping, Opposites, Adding	H.	III	.19	.10 –.14	.28
Tapping, Opposites, Adding	O.	III	–.25	–.20 .17	–.25
Tapping, Opposites, Adding	J.	III	–.35	.20 .15	.30

few other cases when positive and negative results were combined, the corrected coefficient took a position either beyond the largest positive or beyond the largest negative. In all probability, the three tendencies indicate that the employment of the composite measure as one of the variables in the correlation has had the effect of balancing out chance errors.

To be consistent with the procedure so far held to in the discussion of results, that is of considering median coefficients, any general effect of combining tests must be observed by comparing medians.

A glance at the medians in Table XIV and those presented here show that correction does not affect the median coefficients very much. For Strength of Grip the correction for the women gives .00, midway between .13 and –.13 for the right and left hand. For the men it is –.19 increasing the tendency in the negative direction from –.10 and –.06 for the

right and left hand. For the results as a whole, –.07 is in the direction of an intermediate position between .065 and –.095. For the women of Group I and Group II the median for the composite measure for Color Naming and Tapping, .21 .25, respectively, is .10. When the results for Group III (the Opposites Test and Tapping) are included the median coefficient is increased on the negative side. The tendency for the composite measure in both groupings is in the direction of a negative correlation. For the men of Group II in the same tests an intermediate position is maintained .06, between –.06 and .08. When the results for Group III are added the correlation is still more positive, .07 consistent with the tendency indicated by the raw median coefficients .13 and .00. For the men and women combined, excluding Group III the median coefficient is .06. Including Group III it is –.06 in either case somewhat beyond the extremes for the raw coefficients .035 .05.

TABLE XIV

MEDIAN COEFFICIENTS FOR COMPOSITE MEASURES

	Muscular Strength Right and Left Hand Grip	Speed Color Naming and Tapping	Speed Color Naming Addition Av. Reaction Time	Accuracy Color Naming Addition Indiv. Reactions
Women, Groups I & II	.00	.10	.02	.17
Men, Groups I & II	–.19	.06	–.34	–.04
Combined	–.07	.06	–.19	–.08

For Group III the rankings in the Tapping Test, the Opposites Test and the Addition Test were combined to secure a composite measure for speed. The rankings of errors in the Opposites Test, and in Addition were combined with the rankings in the Individual Reactions to secure a composite measure for Accuracy. When the correlation coefficients for this measure of speed are included with the two measures for speed for Groups I and II the medians are as follows:

TABLE XV

	Muscular Strength Right Hand Left Hand Grip	Speed Color Naming and Tapping	Speed Color Naming Add'n. Avg. React. Time	Accuracy Color Nam'g Add'n. Ind. Reaction
Women	.00	–.14	–.21	.17
Men	–.19	–.07	–.04	–.10
Combined	–.07	–.06	–.135	–.08

For the composite measure for speed (Color Naming (.02), Addition Test (.–15) and Average Reaction Time in the Free Association Test (–.195) the median coefficient for the women is .02, exclusive of Group III. Inclusive of Group III it is –.21 intensified in the negative direction due to the additional negative coefficient –.04 for the Opposite Test.

For the men of Group I and II the composite measures give for the same tests a median coefficient of –.34. Including Group III the median is –.04, the raw coefficients being respectively .13, .08 and –.15.

For the composite measure for accuracy (Errors in the Color Naming and Addition Test and Individual Reactions), the median coefficient is slightly lower for the women, .17 instead of .195; for the men it is considerably lower, –.04 instead of –.17; for the results as a whole –.08 instead of –.10, but still consistent with the medians for the raw coefficients.

A further attempt to take account of attenuation was made by adding the original ranking for Pulse and Systolic Blood Pressure and using the sum as a composite measure that might correlate better with mood than did either of the measures taken singly. The same general tendency noticed in the results for the composite measure in the tests of performance appears here. In Group I two corrected coefficients are increased in a negative direction; another is increased in a positive direction with the median at –.32 instead of .05 and –.20, respectively, for the raw median coefficients for Pulse and Blood Pressure. In the case of one man and one woman of Group II, a positive coefficient for Pulse and a negative coefficient for Systolic blood pressure give a low negative corrected coefficient –.02 in each case. In three cases a negative coefficient for Pulse and a positive coefficient for Systolic Blood Pressure result in a positive corrected coefficient; in two cases the corrected coefficient falls between the two raw coefficients and in one at the positive extreme. In two cases when the coefficients for Pulse and Systolic Blood Pressure were negative the corrected coefficient is increased in the negative direction beyond the position of the largest negative. Out of eight cases of negative coefficients for Pulse, four remained negative and four became position: In five out of six cases of negative coefficients for Systolic Blood Pressure the corrected coefficient remained negative. In three out of four cases of positive coefficients for Systolic Blood Pressure the

corrected coefficient remained positive. As was the case with the composite measures used in the tests of performance, the median coefficients here are but slightly affected, being –.03 for the women and –.02 for the men with –.02 for the men and women taken as a group. For the women the corrected median coefficient falls between –.01 and –.05 for Pulse and Blood Pressure respectively; for the men it is nearer zero than either raw coefficients –.04 and .08; for the results combined it falls within the limits of the raw coefficients, –.02 and –.08.

Still another attempt to secure a measure that might be freer from attenuation was made by combining the scores for Pulse and Pulse Pressure. The results appear in Table XII.

TABLE XVI

MEDIANS FOR COMPOSITE MEASURES OF CIRCULATION

	Pulse and Systolic Blood Pressure	Pulse and Pulse Pressure
Group I —Women	–.32	
Group II—Women	–.02	–.06
Combined	–.03	
Group I —Men		
Group II—Men	–.02	–.14
Men and Women Combined	–.02	–.10

For the women four cases of negative coefficient for Pulse remain negative in the combined measure; for the men three out of four remain negative. In the case of one woman the position coefficients for both Pulse and Pulse Pressure remain positive, falling between the position of each. In the case of one man showing a negative coefficient for Pulse and a position coefficient for Pulse Pressure, as a result of combining tests, the coefficient is positive and intermediate between the two.

In another case two positive coefficients a low negative, –.02 appears as the result of combination. As regards Pulse Pressure, for the women, two negative coefficients remain negative, two positive become negative, one positive remains positive. For the men two negative coefficients remain negative, two positive become negative, and one positive remains positive but lower than either of the raw coefficients. For the women the median coefficient for combining Pulse and Pulse Pressure is –.06 as against –.02 for combining Pulse and

Systolic Blood Pressure. For the men it is −.14 somewhat lower than the median −.02 which resulted from combining Pulse and Systolic Blood Pressure.

Although this method of combining tests has had so slight an affect upon the median coefficient, both in the tests of circulation and of performance, the method has served a useful purpose in that it has furnished a sort of check upon the reliability of the work.

CHAPTER IV

SECTION I

CIRCULATION AND PERFORMANCE

It seemed advisable to extend this study into an inquiry of the relation between circulation and mental and physical performance. To that end the scores for Pulse, Systolic Blood Pressure and Pulse Pressure were ranked and each correlated with performance in the several tests in the same manner as with the scores for Mood. This observation is confined to the results for Group I and Group II, the test for circulation having been omitted with Group III because of the impracticability of obtaining such measures under conditions required for group tests. The coefficient for these separate correlations with the measure of their reliability appear in the following three tables.

In all instances the figures for reliability mean that if the experiment were repeated in the same way and under the same conditions what are the chances out of 100 that the median will be equal to or greater than zero.

TABLE XVII

Pulse Rate

	Subjects	Line	Two Point Threshold	Right Hand Grip	Left Hand Grip	Tapping	Color Naming	Errors	Addition	Errors	Individual Reactions	Average Reaction Time
					Women							
Group I	L.	.82	-.72	—	—	1.00	.66	-.21	-.62	-.80	.21	.36
	K.	-.66	.21	.37	—	-.10	-.16	.21	-.36	-.10	.27	.15
	M.	.10	.77	.51	—	.51	.62	-.19	-.35	.33	.51	.38
Group II	L.	-.58	.46	.27	.21	.27	.35	-.42	-.06	.08	-.02	-.44
	S.	.56	—	.27	.21	.21	.19	.03	.25	.06	.06	.02
	K.	.44	.31	-.26	.25	.02	.17	.08	-.15	-.04	.26	.13
	H.	-.23	-.25	.44	.21	.23	-.37	-.35	.10	.10	.27	.33
	P.	-.02	.02	.08	-.02	-.19	-.17	-.08	.02	.25	-.48	-.54
Medians Groups combined		.04	.21	.27	.21	.22	.18	-.135	-.10	.08	.02	.14
Average		.05	.11	.24	.08	.37	.16	-.12	-.14	-.01	.00	.07
S. D.		.50	.46	.24	.17	.37	.35	.20	.27	.32	.31	.32
P. E. of Med.	±	.0151	.1492	.0778	.0652	.1293	.1054	.0602	.0813	.0964	.0726	.1227
Med./P. E.		2.66	1.4	3.2	3.2	2.8	1.7	2.2	1.2	.83	.27	1.14
Reliability		96.02	82.75	98.45	98.45	96.48	87.42	6.89	20.92	72.23	57.37	77.09
					Men							
Group I	C.	-1.00	.27	.52	—	-1.00	-1.00	.14	.52	1.00	1.00	.52
Group II	F.	.21	-.23	-.08	.15	-.19	-.25	.25	.27	.19	-.25	.06
	B.	-.13	-.04	-.17	-.17	.17	.17	-.08	-.13	-.33	.21	-.02
	O.	.17	-.13	-.08	.00	-.19	.33	-.29	.33	.06	-.27	-.06
	T.	.08	—	.21	-.46	-.25	.37	.25	.33	.13	.08	.17
	Km.	.13	.02	.17	.23	-.33	.19	-.42	-.13	.04	-.48	-.06
Medians Group II		.13	-.085	-.08	.00	-.19	.19	-.08	.27	.06	-.25	-.02
Average		.09	.08	.01	.05	.14	.16	-.06	.13	.02	.14	.02
S. D.		.11	.17	.15	.27	.33	.22	.27	.22	.18	.38	.17
P. E. of Med.	±	.0421	.0651	.0575	.1035	.1267	.0843	.1035	.0843	.1073	.1457	.0652
Med./P. E.		3.1	1.3	1.4	.00	1.5	2.13	.77	3.4	.57	.17	.306
Reliability		98.17	19.03	17.25	50.00	15.59	92.17	29.18	98.91	64.97	23.52	41.98
Medians Men and Women Combined		.10	.02	.19	.075	.02	.19	-.08	-.06	.08	-.02	.06
Average		.07	.03	.29	.02	.10	.15	.10	-.04	.00	.05	.05
S. D.		.39	.36	.27	.02	.35	.29	.30	.28	.28	.29	.26
P. E. of Med.	±	.0908	.0863	.0669	.0054	.0819	.0619	.0703	.0656	.0566	.0679	.0539
Med./P. E.		1.1	.24	4.00	1.4	.24	2.89	1.12	.91	1.2	.29	1.1
Reliability		77.09	54.04	99.65	82.75	56.43	97.05	22.91	27.42	75.42	42.48	77.09

TABLE XVIII

Systolic Blood Pressure

	Subjects	Line	Two Point Threshold	Right Hand Grip	Left Hand Grip	Tapping	Color Naming	Errors	Addition	Errors	Individual Reactions	Average Reaction Time
Women												
Group I	L.	.62				1.00						
	M.											
Group II												
		.44				.40		.15	.06	.17	.06	
Medians Group combined		.22	.10	.10	.06	.19	.00	.08	.057	.057	.00	.125
Average		.26	.15	.17	.26	.21	.09	.09	.18	.13	.08	.20
S. D.		.33	.13	.22	.24	.44	.30	.30	.30	.30	.13	.22
P. E. of Med.	±	.1533	.1070	.1362	.0919	.1325	.0902	.0902	.1175	.1180	.0994	.1265
Med./P. E.		1.4	.93	.73	1.43	1.43	.88	1.33	.85	.85	.00	1.45
Reliability		82.75	26.39	31.13	66.85	83.27	50.00	41.20	66.95	66.95	50.00	83.60
Men												
Group I	C.	-1.00				-1.00	-1.00	.14	.52	1.00	1.00	.52
Group II	O.											
	Km.											
Medians Group II		.13	.135			.21	.02	.06	.04	.17	.06	.21
Average		.31	.15	.18	.18	.17	-.12	-.10	.03	.10	.16	.21
S. D.		.43	.13	.23	.27	.20	.31	.28	.15	.32	.39	.23
P. E. of Med.	±	.1226	.0548	.1609	.1039	.0655	.1188	.1073	.0574	.1226	.1878	.0881
Med./P. E.		1.06	2.46	.99	3.17	3.2	.16	.56	.15	1.37	.32	3.5
Reliability		78.30	95.00	62.97	99.37	98.45	54.29	35.28	68.16	16.87	58.54	99.12
Medians Men and Women Combined		.13	.00	.03	.195	-.06	.02	-.06	.04	.00	.02	.295
Average		.08	.02	.04	.17	.06	.01	.08	.12	.05	.08	.14
S. D.		.33	.31	.44	.27	.38	.33	.31	.34	.30	.25	.37
P. E. of Med.	±	.0774	.0772	.1091	.0734	.0659	.0745	.0239	.0726	.0703	.0586	.0965
Med./P. E.		1.68	.00	.27	2.7	27.96	56.96	8.25	64.80	50.00	99.06	98.70
Reliability		87.15	50.00	57.23	96.57	.91	.26	2.08	.55	.00	.34	3.3

TABLE XIX
Pulse Pressure

Group II	Subjects	Line	Two Point Threshold	Right Hand Grip	Left Hand Grip	Tapping	Color Naming	Errors	Addition	Errors	[Individual] Reactions	Average Reaction Time
						Women						
Medians												
Average												
S. D.												
P. E. of Med.												
Med./P. E.												
Reliability												
						Men						
Group II												
Medians												
Average												
S. D.												
P. E. of Med.												
Med./P. E.												
Reliability												
Medians Men and Women Combined.												
Average												
S. D.												
P. E. of Med.												
Med./P. E.												
Reliability												

It is readily observed that the results appearing in the tables are not unlike those secured for mood. The same wide difference is observable with regard to the performance of the separate individuals and with regard to the individual's own performance in the separate tests. Any agreement that might exist may be determined by counting instances of similarity of results. These instances expressed in per cents appear below together with corresponding per cents for Mood between which a comparison can readily be made:

TABLE XX
GROUPS I AND II

Mood	+	−	00	No. of Coefficients	Pulse	+	−	00	No. of Coefficients
Women	50	31	2	83	Women	48	35		83
Per cts.	60	.38	.02		Per cts.	.58	.42		
Men	28	36		64	Men	33	30	1	64
Per cts.	44	.56			Per cts.	.52	.47	.01	
Total					Total				
M. & W.	78	68	2	147	M. & W.	81	65	1	147
	.53	.46	.01			.55	.44	.006	

Systolic Blood Pressure	+	−	00	No. of Coefficients	Pulse Pressure	+	−	00	No. of Coefficients
Women	39	41	3	83	Women	38	13	3	54
Per cts.	.47	.49	.04		Per cts.	.71	.24	.05	
Men	39	23	2	64	Men	31	21	2	54
Per cts	.61	.36	.03		Per cts.	.56	.39	.04	
Total					Total				
M. & W.	78	64	5	147	M. & W.	69	34	5	108
	.53	.44	.03			.64	.32	.04	

Mood. For the women, including Group III, out of 160 correlation coefficients 88 or 55 per cent are positive, 70 or 44 per cent are negative while 2 or 1 per cent are zero. For the men including, Group III, out of 148 correlation coefficients 79 to 53 per cent are positive and 69 or 47 per cent are negative. For the group as a whole out of 308 correlation coefficients 167 or 54 per cent are positive and 139 or 45 per cent are negative and 2 or .006 are zero.

Taking Groups I and II for whom coefficients of correlation for circulation and performance have been worked out also, we observe that out of 83 coefficients 50 or 60 per cent are positive, 31 or 38 per cent are negative and 2 or 2 per cent are zero. For the men of these two groups out of 64 coefficients 28 or 44 per cent are positive and 36 or 56 per cent are negative. For the groups as a whole, out of 147 coefficients

78 or 53 per cent are positive, 68 or 46 per cent are negative and 2 or one per cent are zero.

Pulse. For the women, out of 83 correlation coefficients 48 or 58 per cent are positive and 35 or 42 per cent are negative. For the men, out of 64 correlation coefficients 33 or 52 per cent are positive and 30 or 47 per cent are negative. For the group as a whole, out of 147 correlation coefficients 81 or 55 per cent are positive and 65 or 44 per cent are negative while 1 or less than 1 per cent are zero. For the women there is a slight falling off in positive coefficients as compared with the results for mood; with the men there is a rather noticeable increase in positive correlation for pulse and performance. For the group as a whole the tendency is still toward an increase in positive coefficients but the difference is not considerable, 55 per cent as compared with 53 per cent in the case of Mood.

Systolic Blood Pressure. For the women, out of 83 correlation coefficients, 39 or 47 per cent are positive, 41 or 49 per cent are negative and 3 or 4 per cent are zero. For the men, out of 64 correlation coefficients, 39 or 61 per cent are positive, 23 or 36 per cent are negative and 2 or 3 per cent are zero. There appears a slight tendency toward a negative correlation with systolic blood pressure in the results for the women and a rather noticeable tendency toward a positive correlation in the results for the men. Considering the group as a whole, the positive tendency is less than for pulse, 53 per cent as compared with 55 per cent, but the same as that appearing in the results for Mood.

Pulse Pressure. (Group II only is included in these figures). For the women, out of 54 correlation coefficients, 38 or 71 per cent are positive, 13 or 24 per cent are negative and 3 or 5 per cent are zero. For the men, out of 54 correlation coefficients, 31 or 56 per cent are positive and 21 or 39 per cent are negative, and 2 or 4 per cent are zero. There is a marked increase in the results for both men and women in the number of positive coefficients. This same increase in noticed in the combined results, where out of 108 correlation coefficients 69 or 64 per cent are positive and 34 or 32 per cent are negative and 5 or 4 per cent are zero. This ratio of 2 to 1 in favor of positive coefficients is more marked than in the case of Pulse, Systolic Blood Pressure, or Mood.

The coefficients secured from correlating measures of cir-

culation with scores in the ten tests that are more definitely tests of performance were plotted in the form of a distribution curve. Table XXI gives the results of the distribution —the 25 percentile, the 75 percentile, the median, the P. E. of the median, Q and the range. Table XXII shows the correspondence of the actual distribution to the normal distribution. The range for Pulse Rate and Systolic Blood Pressure is about the same. It is somewhat narrower for Pulse Pressure due to the smaller number of measures represented in the distribution. The medians in all three measures of circulation are within .10 of zero, and the 25 percentile and 75 percentile of all three measures are within .26 of zero. Considering the sexes separately or combined, the actual distribution does not depart sufficiently from the normal distribution to indicate that mood had influenced results.

TABLE XXI

	25 Percentile	Median	P. E. of Median	75 Percentile	Q	Range	No. of Coefficients
PULSE RATE							
Women	−.114	+.083	±.032	+.28	.227	−.80 to +1.00	75
Men	−.178	+.032	±.023	+.221	.199	−1.00 to +1.00	58
Men and Women Combined	−.176	+.06	±.023	+.26	.218	−1.00 to +1.00	133
SYSTOLIC BLOOD PRESSURE							
Women	−.232	−.026	±.037	+.234	.258	−.82 to +.82	75
Men	−.149	+.10	±.048	+.344	.246	−1.00 to +1.00	58
Men and Women Combined	−.205	+.034	±.023	+.321	.263	−1.00 to +1.00	133
PULSE PRESSURE							
Women	−.029	+.08	±.024	+.243	.136	−.42 to +.42	49
Men	−.124	+.07	±.031	+.234	.179	−.56 to +.72	49
Combined	−.06	+.076	±.019	+.24	.15	−.56 to +.72	98

TABLE XXII

	−4Q	−3Q	−2Q	−1Q	+1Q	+2Q	+3Q	+4Q
Normal Distribution		.012	.067	.16	.50	.16	.067	.012	
PULSE RATE								
Women		.04	.04	.173	.583	.12	.066	.0133	.0133
Men	.034		.051	.172	.50	.155	.051		.034
Combined	.0226	.015	.045	.172	.518	.135	.067	.0073	.0226
SYSTOLIC BLOOD PRESSURE								
Women		.013	.04	.186	.50	.11	.066	.04	
Men	.084		.086	.137	.517	.171	.017	.034	
Combined		.022	.045	.187	.555	.097	.067	.022	
PULSE PRESSURE								
Women		.02	.12	.02	.612	.133	.04		
Men		.0204	.1022	.142	.469	.133	.061	.0204	
Combined		.051	.082	.102	.5204	.163	.051	.0102	.0102

Again to be consistent with the procedure adopted for the discussion of results in this study any relation between the various series of correlation coefficients must be observed by comparing medians.

Table XXIII shows the median coefficients in the separate tests for Mood, Pulse, Systolic Blood Pressure, and Pulse Pressure for Groups I and II.

It will be observed that in nine out of eleven tests the women tend toward a positive relation between performance and Pulse and Pulse Pressure, and in four tests toward a positive relationship with Systolic Blood Pressure. The men show a positive relation with Systolic Blood Pressure in nine of the eleven tests, with Pulse in four, with Pulse Pressure in six, and with Mood in two tests out of the eleven. For the men and women combined the tendency to positive correlation is evident in eight of the eleven tests with Mood and Pulse, in nine with Pulse Pressure and in seven with Systolic Blood Pressure.

Table XXIV gives the range of the median coefficients in the various tests with the median coefficients for the four relations.

For the women the range in Pulse and Systolic Blood Pressure is slightly wider than for Mood and about the same for Pulse Pressure. The median for Pulse is slightly higher than for Mood, for Pulse Pressure slightly lower and for Systolic Blood Pressure still lower. The range for the mean

TABLE XXIII

MEDIAN COEFFICIENTS

(Groups I and II)

	Women				Men				Men and Women Combined			
	Mood	Pulse	Systolic Blood Pressure	Pulse Pressure	Mood	Pulse	Systolic Blood Pressure	Pulse Pressure	Mood	Pulse	Systolic Blood Pressure	Pulse Pressure
Line	.21	.04	.22	.27	.04	.13	.13	−.17	.17	.10	.13	.115
Sensitivity	.23	.21	−.10	.125	−.065	−.085	.135	.03	.02	.02	.00	.125
Right Hand Grip	.13	.27	−.10	.06	−.10	−.08	.08	.19	.065	.19	.03	.07
Left Hand Grip	−.13	.21	−.06	.21	−.06	.00	.33	.19	−.095	.075	.195	.20
Tapping	.20	.22	−.19	.04	.08	−.19	.21	.10	.15	.02	−.06	.07
Color	.12	.18	.00	.08	−.06	.19	.02	−.33	.03	.19	.02	.03
Errors	.02	−.135	−.03	.23	−.13	−.08	−.06	.08	.02	−.08	−.06	.15
Adding	−.125	−.10	.075	.02	−.33	.27	.04	−.02	−.17	−.06	.04	.00
Errors	.115	.08	.075	.23	.19	.06	−.17	.00	.15	.08	.00	.08
Indiv. Reaction	.175	.02	.00	.00	−.08	−.28	.06	−.04	.05	−.02	.02	−.02
Avg. React. Time	.195	.14	.185	−.08	−.17	.02	.31	.06	−.10	.06	.295	.06
	9	9	4	9	2	4	9	6	8	8	7	9
	2	2	5	1	9	6	2	4	8	8	2	1
			2	1		1		1			2	1

TABLE XXIV

	Mood Range	Difference	Med.	Pulse Range	Difference	Med.	Systolic Blood Pressure Range	Difference	Med.	Pulse Pressure Range	Difference	Med.
Women	−.13 to .23	36	.13	−.13 to .27	40	.14	−.19 to .22	41	.00	−.08 to .27	.35	.08
Men	−.33 to .19	52	−.04	−.25 to .27	52	.02	−.17 to .33	50	.08	−.33 to .19	.52	.03
Combined	−.17 to .17	34	.05	−.08 to .19	27	.06	−.06 to .295	35	.03	−.02 to .20	.22	.07

in all four relations is wider than for the women and is practically the same for all four relations. The median position for Pulse and Pulse Pressure is higher than it is for Mood, and highest for Systolic Blood Pressure.

CHAPTER IV

SECTION II

MOOD AND WEATHER

In any investigation of mood the influence of the weather on feelings must be considered. In this experiment observations were made of the weather, the winter days being classified roughly as "Bright," "Average," and "Gloomy." As spring approached the days varied more in temperature and humidity making a slight difference in classification necessary. The days for this part of the experiment were classified roughly as "Bright-temperate," "Bright-cold," "Warm" and "Gloomy."

The mood recorded for the various days was located in its place among the scores already ranked as falling in the lower or the upper half of the range. As the difference between adjacent scores was not at all uniform any finer divisions would have to be made arbitrarily and on the whole would be of but slightly more value in revealing relations. In Table XXV are included the result for the subjects in Group II who took the tests from October 12 to Feb. 18. In Table XXVI are included the result for the subjects in Group III who took the tests from Feb. 28 to May 9.

The women of Group II made 100 judgments of mood covering a period of 21 days, seven of which were "Bright," seven "Average," and seven "Gloomy." Of the 31 judgments of mood on bright days 58 per cent were in the upper levels of

TABLE XXV

Mood in Relation to the Weather

Group II

	Women					Men			
	No. of days	Levels of Mood IV-III	II-I	No. of Judgments		No. of days	Levels of Mood IV-III	II-I	No. of Judgments
Bright	7	.58	.42	31	Bright	10	.60	.40	45
Average	7	.49	.51	35	Average	6	.33	.67	27
Gloomy	7	.27	.73	34	Gloomy	4	.53	.47	19
	21			100		20			91

TABLE XXVI
Group III

	Women				Men				Combined			
	No. of days	Levels of Mood IV-III	II-I	No. of Judgments	No. of days	Levels of Mood IV-III	I-II	No. of Judgments	Levels of Mood IV-III	II-I	No. of Judgments
Bright—Temperate	1	.90	.10	10	1	.50	.50	6	.75	.25	16
Bright—Cold	1	.50	.50	6	1	.62	.38	13	.58	.46	19
				16				19			35
Hot	1	.37	.63	8	1	.33	.67	12	.35	.65	20
Warm	1	.37	.63	8	1	.27	.73	11	.32	.68	19
Warm	1	.50	.50	6	1	.50	.50	8	.50	.50	14
Rather Warm	1	.33	.67	9	1	.46	.54	9	.39	.61	18
				31				40			71
Gloomy	1	.54	.46	9							
Slightly Gloomy					1	.45	.55	11	.50	.50	20
Very Gloomy	1	.46	.54	9	1	.54	.46	9	.50	.50	18
				18				20			38

cheerfulness and 42 per cent in the lower levels. Of the 35 judgments made on "Average" days, 49 per cent were in the upper levels and 51 per cent in the lower levels. Of the 34 judgments made on "Gloomy" days, 27 per cent were in the upper levels and 73 per cent in the lower levels of cheerfulness. The men of Group II made 91 judgments of mood covering a period of 20 days, ten of which were "Bright," six "Average" and four "Gloomy." Of the 45 judgments made on bright days 60 per cent were in the upper levels and 40 per cent in the lower levels of cheerfulness. Of the 27 judgments made on "Average" days, 33 per cent were in the upper levels and 67 per cent in the lower levels. Of the 19 judgments made on "Gloomy" days, 53 per cent were in the upper levels and 47 per cent in the lower levels of cheerfulness. The tendency is for both men and women to feel more cheerful on "Bright" days and less cheerful on "Average" and "Gloomy" days. The women seem to be more affected by gloomy weather than the men, who indicate the greatest influence on "Average" days. Possibly this may be explained by the fact that the "Average" days of this particular winter were seldom free from some of the conditions that prevailed on the "Gloomy" days.

The women of Group III made 16 judgments of mood, ten being made on a "Bright-temperate" day and six being made on a "Bright-cold" day. On the first, 90 per cent were in the upper levels of cheerfulness and 10 per cent in the lower levels. On the second, 50 per cent were in the upper levels and 50 per cent in the lower of cheerfulness. For the men and women combined of the 16 judgments made on the Bright-temperate day, 75 per cent were in the upper levels, and 25 in the lower levels. Of the 19 judgments made on the "Bright-cold" day 58 per cent were in the upper levels and 46 per cent in the lower levels of mood.

The opposite relation exists for the four "Warm" days. Of the 71 judgments made on three of the warm days, the smallest per cents are in the upper levels, the per cent in the lower levels of cheerfulness being more than one and one-half as great as in the upper levels. On one of these days there is no difference in the per cents for either women or men.

On the slightly "Gloomy" days the women of this group gave more judgments in the upper levels of cheerfulness, the per cent being 54, as opposed to 46 per cent in the lower. On

the very "Gloomy" day, the relation is reversed, 46 per cent of the judgments being in the upper levels and 54 per cent in the lower levels. The men of this group as was the case with the men of group II tend to be lower in mood on slightly gloomy days than they are on definitely gloomy days. In the first instance, 45 per cent of the judgments are in the upper levels as against 55 per cent in the lower and in the instance, 54 per cent are in the upper as against 46 per cent in the lower levels of mood. Considering the men and women of this group combined the judgments are equally divided between the upper and lower levels of cheerfulness.

Concluding then from the data here presented there appears to be a rather definite relation between mood and weather, a cheerful mood tending to prevail on bright, pleasant days and various degrees of depression on gloomy, disagreeable days.

These results are in harmony with those found by Dexter in an empirical study of the mental and physiological effects of meteorological conditions. According to this investigator emotional states are plainly influenced by the weather states, "hot, humid, cloudy, wet days," although making us "feel out of sorts and liable to be troublesome, show the least number of misdemeanors owing to the fact that the energy is lacking to carry out the emotion because of the depleting effect of the weather."

BIBLIOGRAPHY

1. Ach, N. — Uber den Willensakt und das Temperament.
2. Baldwin. — Dictionary of Philosophy and Psychology.
3. Cannon. — Bodily Changes in Pain, Fear and Rage.
4. Cattel, J. McK. — A Statistical Study of American Men of Science. Science N. S. Vols. 24 and 32.
5. Cobb, Margaret. — Archives of Psychology No. 50.
6. Cogan, L. C., Conklin A., and Hollingworth, H. L. — An Experimental Study of Self-Analysis. School and Society, 1915. 2. P. 171.
7. Crile. — The origin and Nature of the Emotions.
8. Davenport, C. B. — Heredity of Constitutional Mental Disorders. Psych. Bull., 1920.
9. Dexter, Edwin, G. — Conduct and the Weather. Psych. Rev. Mon. Supp. No. 10, 1899.
10. Gley. — Internal Secretions.
11. Hollingworth, H. L. — Experimental Studies in Judgment. Arch. of Psych. No. 29.
12. Hollingworth, H. L. — The Influence of Caffein on Mental and Motor Efficiency. Arch. of Psych. No. 22. P. 13.
13. Hollingworth, H. L. — Correlation of Abilities as Affected by Practice. Jr. of Educ. Psych. Sept. 1913.
14. Hollingworth, L. S. — Functional Periodicity. Teachers College, Columbia University Contributions to Education, No. 69.
15. James. — Principles of Psychology.
16. Janeway. — Lecture Notes on Physiology—The Ductless Glands.
17. Kent-Resanoff. — A Study of Association in Insanity. Amer. Jr. of Insanity. LXVII.
18. McPhail, A. H. — Adenoids and Tonsils: A Study Showing how the Removal of Enlarged or Diseased Tonsils Affects a Child's Work in School. Ped. Sem. June, 1920. Pp. 188-194.
19. Morgan, J. J. B. — The Overcoming of Resistances and Other Distractions. Arch. of Psych. No. 35.
20. Pressey & Chambers. — A. Scale for Investigating the Emotions. Jr. of Appl. Psych. 1920. 4: 92-104.
21. Poffenberger, A. T. — Influence of Practice on Mental Processes. Jr. of Educ. Psych. Vol. 6. 1915, P. 462.
22. Rugg, H. O. — Is the Rating of Human Character Practicable? Jr. of Educ. Psych. Nov., Dec., 1921, Jan., Feb. 1922.
23. Sherrington. — The Integrative Action of the Nervous System.
24. Stecher, Lorle I. — The Effect of Humidity on General Efficiency. Arch. of Psych. No. 38 P. 12.
25. Strong, E. K., Jr. — Effect of Hookworm Disease on the Mental and Physical Development of Children. Publ. of the International Health Commission. The Rockefeller Foundation, New York. 1916. 3. P. 121.
26. Thorndike, E. L. — Mental and Social Measurements. P. 194.
27. Thorndike, E. L., McCall W. A., and Chapman, J. C. — Ventilation in Relation to Mental Work.
28. Thorndike, E. L. — Mental Fatigue. Psych. Rev. 1900. Vol. 7., P. 481.
29. Wallin. — The Mental Health of the School Child. Chap. XIII.

30.	Webb, E.	Character and Intelligence. An Attempt at an Exact Study of Character. British Jr. of Psych. Mon. Suppl. Vol. I No. 3.
31.	Woodworth.	Psychology. Chap. XXI.
32.	Woodworth and Wells.	Association Tests. Psych. Mon. No. 57.
33.	Woodworth.	The Accuracy of Voluntary Movement. Psych. Rev. Suppl. No. 13.
34.	The Scott Company Laboratory.	Reprint from the Journal of Applied Psych. Vol. IV. 1920. Pp. 115-125.

THE INFLUENCE OF INCENTIVE AND PUNISHMENT UPON REACTION-TIME

BY

ALBERT M. JOHANSON, Ph.D.

ARCHIVES OF PSYCHOLOGY
R. S. WOODWORTH, EDITOR

No. 54

NEW YORK
MAY, 1922

CONTENTS

Introduction	3
Chapter I. Facilitating Factors	5
1. History of the Inquiry into Suggestion	5
2. History of the Inquiry into Incentive	10
Chapter II. The Experiment	17
1. The Apparatus	17
2. The Procedure	21
3. Standard Conditions	23
Chapter III. The Results	25
1. Reliability of the Measures	34
Chapter IV. Conclusion	50
Chapter V. General Summary	53

INTRODUCTION

Many of the earlier investigators of reaction-time conducted experiments in which such factors as the personal equation, the physiological time of conduction, the kind of stimulus, its intensity, duration and so forth, the optimal interval between the foresignal and the stimulus were the matters of most concern. A later tendency seemed to be towards a consideration of the internal conditions attendant upon the reaction. Such factors as the direction of the attention, practice and fatigue, and the reaction movement were seriously studied. More recently attention was directed to the introspective analyses of the reaction. By this method such problems as those of the reaction movement and expectant attention in reactions were investigated.

The present investigation was conducted with a view towards considering the external conditions and their results upon the reaction. The effect upon the time of the reaction was studied by varying these conditions and noting the variations in the results. A comparison with the conditions recognized as "standard" was thus possible. Little if anything has as yet been done along these lines of varying the accompanying factors, so the history of previous works will perforce be brief.

This investigation may be divided into two parts. The first is a study of the influence of incentive upon the normal reaction to sound. The second part comprises a consideration of the influence of punishment upon the normal reaction to sound.

The writer desires to express his great obligations to Professor Woodworth, Professor Hollingworth, and Professor Poffenberger, for their kind interest and guidance, and to Mrs. Helen H. Wheeler, Mr. Joseph L. Holmes and Mr. Howard K. Nixon, of the Department of Psychology, for their generous assistance as subjects in this experiment.

CLEMSON LIBRARY

Chapter I

FACILITATING FACTORS

Though not a factor in the investigation, suggestion is so closely akin to incentive, as it is used here, that it seemed it would be of interest and would give a setting to the problem to have it included. Some of the investigations, though under the name of suggestion, are not strictly confined to suggestion, but include also an element of incentive which is of primary interest and ought, therefore, to be mentioned. Furthermore, the effects of incentive are also the effects in some forms of the suggestion experiments, and as such are worthy of review. But, since the question of suggestion is only subsidiary to our real problem, it will not be necessary to give a complete account of its rather extensive literature. Only the more noteworthy and those bordering on incentive will be mentioned.

1. History of the Inquiry into Suggestion.

The field of suggestion is replete with experimental investigations of the bearing of suggestion upon various types of work and reactions to stimuli as met with under everyday conditions; upon the great range of individual differences in responding to suggestion, but as yet little has been accomplished in the way of investigating the part played by suggestion under rigidly standardized conditions when measuring the time of a response to a stimulus.

One of the earliest to report his work on suggestion was Binet[1]. He reports an experiment on children in which the subject was to draw a line equal in length to a line shown them. In these reproductions there was a marked tendency to draw the line too short. Seventy-eight out of eighty-six made it less than fifty millimeters, the length of the original. The reproductions varied in length from twenty-eight to sixty millimeters. They were then informed that a longer line would be shown, which they were to reproduce in like manner. The line which they were shown, however, was actually shorter, measuring only forty millimeters. Only nine of the eighty-six drew the second line shorter than the first, and of these only one drew it ten millimeters shorter. They were then told that a third line a little shorter than the second would be shown them, whereupon the original line of fifty millimeters was presented again. This sug-

[1] Binet, "La Suggestibilite", L'annee Psychol. V., 99, 1899.

gestion was less efficacious than the first; only seventy of the eighty-six yielding to it. The amount of error was likewise reduced.

Pearce[2] tested the influence of a suggestive stimulus upon the extent of eye movements as indicated by visual localization. He noted that with a single peripheral stimulus the error toward the fixation point increased with the distance of the peripheral stimulus from that point. With the suggestion stimulus there was at first a tendency to resist the suggestion, but this diminished as the suggestion was repeated. The resistance was most vigorous when the suggestion was contrary to the normal error; but ultimately the suggestion opposed to the normal tendency was most effective. Contrary to later investigators, Pearce found that the same person showed the highest degree of suggestibility in all tests whether with visual, auditory or tactual stimuli.

Smith and Sowton[3], though not the next in point of time to report their investigations, are mentioned at this time as they employed the same method as their predecessors. They have investigated the effect of what they called "successive contrast." A line varying from two to twenty centimeters was shown and immediately afterward a standard line of ten centimeters was exposed. The length of the standard was to be marked off on a line already drawn on a sheet of paper. In all cases, both with and without the modifiers, the standard was underestimated, but for one subject modifiers of two to ten centimeters produced an average increase of 1.2 mm., in the estimation, while modifiers of ten to twenty centimeters caused an average decrease of 0.9 mm. That is, the shorter lines acted as a positive suggestion, the longer ones as negative.

The experiments thus far reported were made with but slight differences in the method of investigation. In the earlier experiments the observer was given no instructions to resist any influence by suggestion. With Brand[4] a radically new method was adopted. The subjects were aware of the purpose of the experiment and while the idea of suggestion was to be given place in the mind the observers were warned against any voluntary response to it. Furthermore, the reactions depended not alone upon visual perception, but primarily upon the power to reproduce. The purpose of his experiment, as he states it, was, "to find out how far and in what direction the visual estimation of a linear magnitude could be affected by suggestion of certain possible errors in such estimation, the subject knowing that

[2] Pearce, "Normal Motor Suggestibility", Psychol. Rev. IX., 1902, 348.

[3] Smith and Sowton, "Observations on Spatial Contrast and Confluence in Visual Perception", Brit. J. of Psych. II., 196-219.

[4] Brand, "The Effect of Verbal Suggestion upon the Estimation of Linear Magnitudes", Psych. Rev. 12, 1905, 41-49.

the suggestions were purely arbitrary, i.e., that they had no reference to any foreseen tendency to err in any direction." He tested the effect of printed mottoes on the reproduction of horizontal lines. These lines varied from twelve to thirty-four centimeters in length and were represented by the interval between two pegs situated at a distance of 120 cm. from the eyes of the observer. The observer reproduced this interval by spacing two similar pegs on a ledge 40 cm. from his eyes. The eight different suggestions used were printed upon white cardboard in letters 1.2 cm. high. The experimenter presented them to the subject by displaying the cardboard for a moment. After the suggestion had been presented, the experimenter displayed two small objects upon his frame and then called upon the subject to respond by setting up his two similar objects upon his own frame at a distance from each other approximating as nearly as possible that of the original objects. It does not seem possible to make very clear-cut deductions from his results, with the possible exception that the two brief suggestions, "make short" and "make long" tend more than the other suggestions to make the reproduced distance greater in magnitude, and the two suggestions, "Don't make too long" and "Don't make too short," tend to a less degree to have the same effect.

Bell[5] conducted an experiment in which two types of suggestion were used,—auditory and visual. For the former suggestion "high" and "low" were used, being spoken by the experimenter just before the presentation of the object to be reacted upon. For the visual suggestion a diamond-shaped figure twenty centimeters long and four centimeters wide was shown to the subject. The instructions were to reproduce the triangles of different shapes and heights as they were presented. Three were reproduced with "high" suggestions, and three with "low" suggestions, and three without any suggestions.

Bell concludes that in general the suggestions do affect the reproduction of the triangles; that the auditory suggestion is more effective than the visual; and that in the auditory set the "low" suggestion is more effective than the "high." Throughout the experiment susceptibility to "low" suggestion was more general and more uniform than to the "high" suggestion. But the weakness of his experiment lies in the small number of observations made for each subject with each type of suggestion. Only three observations were taken for each type. Averages drawn from so few cases are of very doubtful reliability.

[5] Bell, "The Effects of Suggestion upon the Reproduction of Triangles and of Point Distances", Amer. J. of Psych. XIX., 1908, 504.

Strong[6] made a more creditable investigation. His experiment was to give the subject a suggestion and have him respond each time with his maximum grip. Collin's elliptical form dynamometer was used and from it an expression in kilograms was obtained of his muscular activity. In all, seven suggestions were employed which may be classified as follows:

Auditory Suggestion:

Positive, "Now you can make it stronger than usual."

Negative, "Now you can't make it as strong as usual."

Visual Suggestion:

Positive: A plus sign on a card was presented.

Negative: A minus sign on a card was presented.

At the beginning of the experiment each subject was informed that the plus sign was meant to suggest to him that he could make his grip stronger than usual and the minus sign was meant to suggest the contrary. These signs were consequently visual suggestions depending on previous vocal instruction.

Auto-suggestion:

Here the experimenter announced, "Now you can make your own suggestion." The subject understood by this that he was at liberty to suggest to himself either the positive or negative suggestion and to designate his choice to the experimenter by audibly announcing it. In this case as soon as the subject had announced his suggestion the dynamometer was handed to him and the experiment continued as usual.

Neutral Suggestion:

This consisted of an announcement, "Now neutral," and was intended to act merely as a check and guide to what would be the exertion if no suggestion of any sort were given.

Strong concludes that suggestion as a whole heightens the maxima. Negative suggestions tend more than the positive to heighten the maxima, but with some subjects the positive suggestions as a general rule are superior to negative suggestions in this respect. Auto-suggestions tend most strongly of all types of suggestion to heighten the maxima. For some of his subjects, visual was superior to auditory suggestion.

Jones[7] reported an experiment which consisted in having the subject

[6] Strong, "The Effect of Various Types of Suggestion upon Muscular Activity", Psych. Rev. 17, 1910, 279-293.

[7] Jones, G. M., "Experiment on Distance as Influenced by Suggestion of Ability and Inability", Psych. Rev., 1910, 17, 269-78.

place two pegs the same distance apart as two which the experimenter had exposed to his view shortly before. There were six varieties of suggestion employed and response was also made to one signal where no suggestion was offered. The visual suggestion was made by means of printed mottoes, such as, "You are now able," and "You are now unable." The vocal was made by the experimenter to the observer in the same words, and the "auto," was made by the subject to himself in the words, "I am now able," or "I am now unable."

In view of the fact that the judgments made under suggestion, whether affirmative or negative, show so frequently an increase in variability and error beyond that in the judgments made without suggestion, it may be inferred that suggestion does in itself and apart from the actual contents of the suggestion, effect some change in the reproduction of distance.

It is also significant that the suggestion acts to a considerable degree in a direction corresponding to the actual contents of the suggestion given. That is, the error and variability under suggestion of ability were almost always less than when suggestion of inability were made.

The conclusion that Jones draws is that negative suggestions (suggestion of inability) were about twice as effective as were the positive or suggestions of ability. Moreover, the auditory negative suggestions were most effective of all.

Powelson and Washburn[8] in 1913 reported a study to determine the effect of verbal suggestion on judgments, of the affective value of colors. All of the ninety Bradley colored papers were used and the subjects were instructed to judge the pleasantness or unpleasantness of the color. The middle eighteen colors of the series were presented with an accompanying verbal suggestion as to their affective value. In the first sitting for half of the observers, the verbal suggestions accompanying this part of the series were suggestions of unpleasantness, and in the second sitting the suggestions were of pleasantness. The suggestion took the form of favorable or unfavorable adjectives pronounced by the experimenter as the color was shown. For instance, "faded", "delicate", "warm", "crude", were the terms used. Of thirty-five observers, twenty-five gave results indicating a positive effect of suggestion in altering the judgments of affective value. The investigators conclude, therefore, "that direct verbal suggestion regarding the pleasantness or unpleasantness of a color has a fairly decided positive effect upon the judgments of observers of the type and under the conditions found in our investigation."

[8] Powelson and Washburn, "The Effect of Verbal Suggestion on Judgment of the Affective Value of Colors", Am. J. of Psych., 1913, 24, 267-9.

Poffenberger[9] has reported in a Master's thesis some differences in suggestibility found by using an electric shock as the stimulus in a reaction experiment. What he did was to have the subject react as soon as he felt the electricity in his reaction key. The shock stimulus was strong at first, then gradually reduced to a minimal point. The reaction was made just the same. In this part of the experiment the sight of the operator's key being thrown was the suggestion factor.

A second group was tested under slightly varied conditions. The suggestion this time was the sound of a falling screen and the presence of a 4 c.p. light. The lamp was arranged in connection with the falling screen so that the light flashed on one one-hundredth of a second before the stimulus was felt in the key. The results obtained show the suggestion time to be slightly quicker or shorter than the normal reaction-time. He concludes also that there is a wide range of individual differences in suggestion to an electric shock. The most suggestible subject had the shortest suggestion time. And finally, women are more suggestible than men, having a greater proportion of successful suggestion reactions and shorter suggestion-time.

The experiments and investigations thus far reported appear to be the more important ones on suggestion and the ones most closely allied to the present investigation. But digressing a few steps, another field may be entered which, though closely related to suggestion, has a direct bearing upon the problem under discussion.

2. History of the Inquiry into Incentive.

In the preceding section experimental studies directly involving suggestion were reviewed. The results of these several investigations may be generalized in some such statement as: suggestion tends in almost every case to assist in the performance of the task by materially reducing the time of the performance, or by increasing the amount of work accomplished in a stated period of time, but with some individual differences. In this next section, the aim will be to show how other factors, aside from pure suggestion, may affect the performing of a task. This group or set of conditions has been quite generally designated by the term "dynamogenic factors," or again as "facilitating factors."

James[1], writing of the principle of dynamogeny, says, "Every sensorial stimulus not only sends a special discharge into particular

[9] Poffenberger, "Individual Differences in Suggestibility with the Use of an Electric Shock", Master's Thesis, 1910.

[1] James, "Principles of Psychology", I., 90ff., 1890.

muscles dependent on the special nature of the stimulus in question, but it innervates the muscles generally."

Féré[2] has given experimental proof of this. The strength of contraction of the subject's hand was measured by a dynamometer. The maximum strength under simple experimental conditions remains the same from day to day. But, when a sensorial stimulation was given simultaneously with the pressure on the dynamometer, a greater force of contraction was produced than when the sensorial stimulation was absent. The effects of sound and musical notes on the muscular effort were also studied. He found that increasing the loudness and pitch caused corresponding increases of muscular effort. The contraction power was decreased by sad notes, but increased with gay notes. The increase of effort varied with the color of the light used. Red was found to be the most exciting color. Tastes and odors were likewise found to have a dynamogenic effect. Cutaneous sensations—heat, cold and so forth—had a facilitating influence. Féré reported also that more work could be done with the eyes open than closed. In working a Mosso ergograph he found that moving the legs and counting aloud reenforced the fingers in their contractions.

Arps[3] undertook to investigate to what extent awareness and partial awareness of results were factors conditioning efficiency. The question to which he experimentally found an answer was whether a condition of relatively complete awareness of results was more or less favorable to efficiency than was a condition of partial awareness. To what extent did knowledge of the results further efficiency; or to what extent did a lack of knowledge curtail efficiency, were other points of interest in his experiment. Further, was a response in which a knowledge of the results constituted the essential features, more or less efficient, than a response when such knowledge was relatively lacking? The work was done with the ergograph and continued to exhaustion. His conclusions were that the organism functions more effectively when the ergographic tracings were immediately present stimuli. Both the absolute amount of work and the rate of work done under conditions of knowledge of the results exceed that done under conditions of ignorance of results.

Whipple[4], in determining the influence of forced respiration found that it had a reenforcing power on most activities. The tests he used were both physical and mental, including the dynamometer grip,

[2] Féré, "Sensation and Movement", 26-50.

[3] Arps, G. F., "Work with Knowledge of Results versus Work without Knowledge of Results", Psychol. Monog., 1920, 28.

[4] Whipple, "The Influence of Forced Respiration on Psychical and Physical Activity", Am. J. of Psych., 1897, 9, 560.

adding of digits, card sorting, simple reaction to sound, memory span, and several others. He concludes that there is an improvement of the muscular mechanism at the expense of mechanisms of control and of the higher functions generally.

Judd[5] has performed an investigation which has an aspect similar to a part of the present experiment. His work was on "Practice without Knowledge of the Results," from which the conclusion is drawn that greater efficiency results when the observer is aware of the conditions underlying the procedure of the experiment. The several series of reactions in the present work provide an analogous case to that of Judd's.

In his experiment, however, the subject was shown a line drawn so as to slope upward or downward at varying degrees. The subject was to continue the line, doing it without seeing what he was doing. "When a given line appeared the subject was required to place a pencil on the opposite, or unseen side of the screen, in what he regarded as the exact continuation of the line." The screen prevented him from seeing how accurately he had located the point.

The striking fact, which appears in the results of these experiments, is that practice brings little if any change. The first and last day differ from each other about as did the first and second. There is no motive for improvement. The subject cannot see his results and cannot, therefore, judge of their success or failure. Under the conditions which obtained in his experiment, the incentive for improvement was entirely suppressed largely because the subject did not know what was occurring.

Some more recent investigations have been conducted by Chapman and Feder[6]. These men wished to determine to what extent an increased interest stimulated an increased effort. The general method of the experiment was to give extended practice in three tests to two similar groups of children, one group working under normal conditions while the other was motivated by external incentives. The results showed that the incentive exerted a considerable effect on the amount of the product. The incentives used were the individual's results of the previous day published in some conspicuous place.

Another experiment carried out with children as subjects was performed by Bronner[7]. She determined the effect of attitude upon the performance of tests by testing children under different conditions

[5] Judd, "Practice without Knowledge of Results", Psych. Rev. Monog. Supp. II., 1905.

[6] Chapman and Feder, "The Effect of External Incentives on Improvement", J. of Educ. Psych., 1917, 8, 469-74.

[7] Bronner, A. F., "Attitude as it Affects Performance of Tests", Psych. Rev., 1916, 23. 303.

or attitudes towards the tests. By appeals and by reasoning with them and by informing them as to the meaning and value of a good score, many cases with high scores were obtained. Summarizing the viewpoint of the report, she states in part that, "mental attitude has been discussed so far as a factor in the learning process. Thorndike has enumerated the laws of learning. He says, 'Purposive behavior is the most important case of the influence of attitude.' Again, 'It is a general law of behavior that the response to any external situation is dependent upon the condition of the man as well as upon the nature of the situation.' Ruger states that the attitude of confidence was a great aid in the successful solving of puzzles."

Some interesting experiments have been worked out recently by Wright[8] to determine the effect of incentives upon one's work. These experiments, consisting of three series, were conducted for the purpose of comparing quantitatively the amounts of work that were accomplished by the subject working under two different mental attitudes. One, that of mere doing because the subject was told to work as hard as he could and as long as he could with no idea of securing any specified result. The other, that of doing a prescribed task as long as strength endured. With the first task all incentive, as, the watching of the instrument or the keeping track of his progress by counting the strokes was denied the subject; whereas under the second condition the subject was not only permitted to watch his strokes, but was also stimulated to action part of the time by his being requested to count his strokes.

To furnish a definite motive for the second class of responses, blocks were inserted under the carriage of the ergograph. The subject was required to push merely to the block and to exert himself to see how many times he could reach it.

The results show a gain for all the subjects in the work performed under the conditions of the second class, that is, with incentives. Wright interprets his results as follows: "The difference in the mental attitudes of the subject under the different conditions imposed upon him in the performance of his tasks affected in no uncertain manner the results accomplished by him; or, as a more general deduction, in seeking the greatest results in the amounts of work to be secured by bodily exertion, the mental attitude of the subject towards his work should be taken into consideration." It was also shown that when working with the incentive there was less fatigue though more work was accomplished.

[8] Wright, W. R., "Some Effects of Incentives on Work and Fatigue", Psychol. Rev., 13, 1906, 23-34.

In another series, a somewhat different incentive was used. A line was drawn on the recording smoked sheet not quite beyond the reach of the subject. In working with this incentive the subject was instructed to watch his work, count his strokes, put forth his utmost effort with each stroke, and, to endeavor to reach the line as often as possible. In a final series the incentive (a line on the smoked sheet) was completely beyond the reach of the subject.

The conclusions arrived at from this experiment are, that the subject accomplished more work when working under the mental stimulus of having a set task to be performed, than he did when working without a definite aim. In the third part, where the line was beyond reach, a known impossibility to accomplish the required conditions tended to decrease the subject's total results. The feeling of fatigue accompanying work was not so great when the subject was working under the direct stimulus of a definite aim notwithstanding the fact that he had at the time produced an increase in his amount of work.

Cleghorn[9] conducted an experiment to show how voluntary muscular contractions could be reinforced. The subject contracted against a weight of two kilogrammes using the ergograph apparatus. A light flashed into the eye, a sudden sound, and induction shocks applied to the skin were the stimuli used. It was found that a sensory stimulus applied just as the muscle was beginning to contract, caused an increase in the height of the contraction.

Many experiments have been performed in the case of the various natural reflexes showing the facilitating effect. Lombard[10], experimenting upon the patellar reflex, found that when other sensations came in simultaneously with the tap, the jerk was increased. For example, clinching the teeth just before the patella was struck reenforced the knee-jerk. He also studied the effects of irritation of the skin, of exciting the attention, of exciting mental work, of music, and found under these conditions a reenforcing effect.

Yerkes[11], in testing the effect of auditory on visual and tactual stimuli in frogs, found that if the auditory stimulus preceded another stimulus by various time-intervals, it had an alternating reenforcing and inhibitory effect. A reflex movement of the leg was chosen as an indication of the action of the stimuli and the influence of sounds was observed. The auditory stimulus was either the sound of a quick

[9] Cleghorn, A., "Reinforcement of Voluntary Muscular Contractions", Am. J. of Psysiol., 1898, I., 336-45.

[10] Lombard, "Reenforcement of Knee-Jerk", Am. J. of Psychol., 1887, I., 18.

[11] Yerkes, "The Sense of Hearing in Frogs", J. of Comp. Neurol. and Psychol., 15, 4, 1905.

hammer blow or the ringing of an electric bell for a certain interval. The tactual stimulus was given by a rubber point. Reactions to the stimuli were taken in pairs regularly at half minute intervals, first a reaction to the tactual stimulus alone, then a reaction to the same intensity of touch when accompanied or preceded by an auditory stimulus. The influence of the sound is discovered by direct comparison of the tactual reaction of each pair with its corresponding auditory-tactual reaction. He concludes that when the tactual reaction is the greater, the sound has partially inhibited the reaction; when it is the smaller, that it has reinforced the reaction; when the two are equal, that it has been without influence. The tactual stimulus regularly causes a reflex movement of the leg. The sound, on the contrary, never causes the slightest movement.

Hofbauer[12], in a similar experiment on human subjects, arrived at the same conclusions. He found that firing a pistol caused a rise in the ergographic record.

Turley[13] tested the effect of a stimulus on its duplicate in a succeeding series of stimuli. Exactly determinable time-intervals between the stimulus and its duplicate were introduced. He concluded that, "(1) If a stimulus precedes at various time-intervals its duplicate in a series of stimuli, it will alternately inhibit and reenforce the perceiving of the duplicate stimulus. (2) Within four and one-half seconds there are at least three points each of maximum inhibition and maximum reenforcement." "(4) Up to 4.5 seconds, as the time-interval increases, the maximum inhibition generally decreases, while the maximum enhancement correspondingly increases."

Bliss[14] found that the reaction-times to sound heard in two ears were shorter than the reaction-times for the same sound heard in one ear. Poffenberger[15] found that the reaction-times to a light seen with the two eyes were less than the reaction-times to the same light seen with one eye. Todd[16] found that whenever another stimulus was added to a given stimulus the reaction-time was reduced. Or if a third stimulus was added to a pair of stimuli the reaction-time was reduced.

Competition is a strong dynamogenic agent. Experimental proof of this was obtained by Triplett[17], in a laboratory race. Two fishing

[12] Hofbauer, "Interferenz zwischen verschiedenen Impulsen im Centralnervensystem", Pflugers Archives, Bd., 68, 564, 1897.

[13] Turley, "Inhibition and Reenforcement", Harv. Psych. Stud., 2, 1906.

[14] Bliss, "Investigations in Reaction-time and Attention", Yale Studies, 1-4, 31, 1892-6.

[15] Poffenberger, "Reaction-time to Retinal Stimulation", Archives of Psychol., 25, 1912.

[16] Todd, "Reaction to Multiple Stimuli", Archives of Psychology, 25, 1912.

[17] Triplett, "The Dynamogenic Factors in Pacemaking and Competition", Am. J. of Psychol., 1897, 9, 507-33.

reels were used in the experiment. The subject, instructed to turn his reel at the highest possible rate of speed, worked at the side of a competitor who reeled at the same time with a similar instrument. From this simple experiment Triplett concluded as follows: "From the above facts regarding laboratory races we infer that the bodily presence of another contestant participating simultaneously in the race serves to liberate latent energy not ordinarily available. This inference is further justified by the difference in time between paced competition races and the paced races against time, amounting to an average of 5.15 seconds per mile up to 25 miles." "The sight of the movements of the pace makers or leading competitors, and the idea of higher speed, furnished by this or some other means are probably in themselves dynamogenic factors of some consequence."

Part of the dynamogenic effect of competition comes from the social element, and part from each performer's having a definite measure of his performance. Just as in the practice experiment, it has been found that presenting the records to the individual and allowing him to have them before him in the form of a curve, and urging him to compete against his own previous performance, causes the individual to accomplish more and progress faster than he otherwise would, so in this reaction experiment, the incentive of his previous record caused a greater speed to be developed.

CHAPTER II.

THE EXPERIMENT

In all of the foregoing reports of investigations, it has been noted how the factor of incentive helped to produce better results, either by a reduction in the time of the performance, or by an increase in the amount of work accomplished. It has also been noted that experiments with this purpose have been conducted over a very extended period, but no one up to the present time has used either a positive or a negative incentive with reaction-time in the manner in which it has been employed here.

This investigation will show how reaction-time can be reduced by accompanying factors so that a normal time, apparently as quick as the individual is able to respond, is brought much below the usual level of response. In the first part of the experiment an incentive was included to speed up the reaction process. The incentive was the time the subject had taken in his previous response to the stimulus, measured in sigma. The second part of the experiment included the series of reactions accompanied by a negative incentive or punishment. In this case the subject was given an electric shock for slow reactions. The results show a very decided improvement due to this unpleasant conditioning factor.

1. *The Apparatus.*

To record the reaction times, we were fortunate in being able to use the Hipp chronoscope modified and tested by Poffenberger and Morgan[1]. Their principal change was the replacing of the springs by a counterweight as devised by Dunlap[2]. The arrangement of the circuits was according to Dunlap's method. The University current (direct current), reduced through a rheostat so that the strength of the current was four volts, was the main source of supply. The various circuits were led off from the rheostat. One line led to the lower electromagnets of the chronoscope by way of the subject's reaction key. The ordinary telegraph key was used as a reaction key. From the chronoscope the circuit was continued to the upper contact post of the Wundt sound hammer, then through a mercury cup at

[1] Poffenberger and Morgan, "The Hipp Chronoscope: Its Use and Adjustment", J. of Exp. Psychol. I., 3, 1916, 185-199.

[2] Dunlap, K., "The Hipp Chronoscope without Armature Springs", Brit. J. of Psychol., 1912, 5, 1-7.

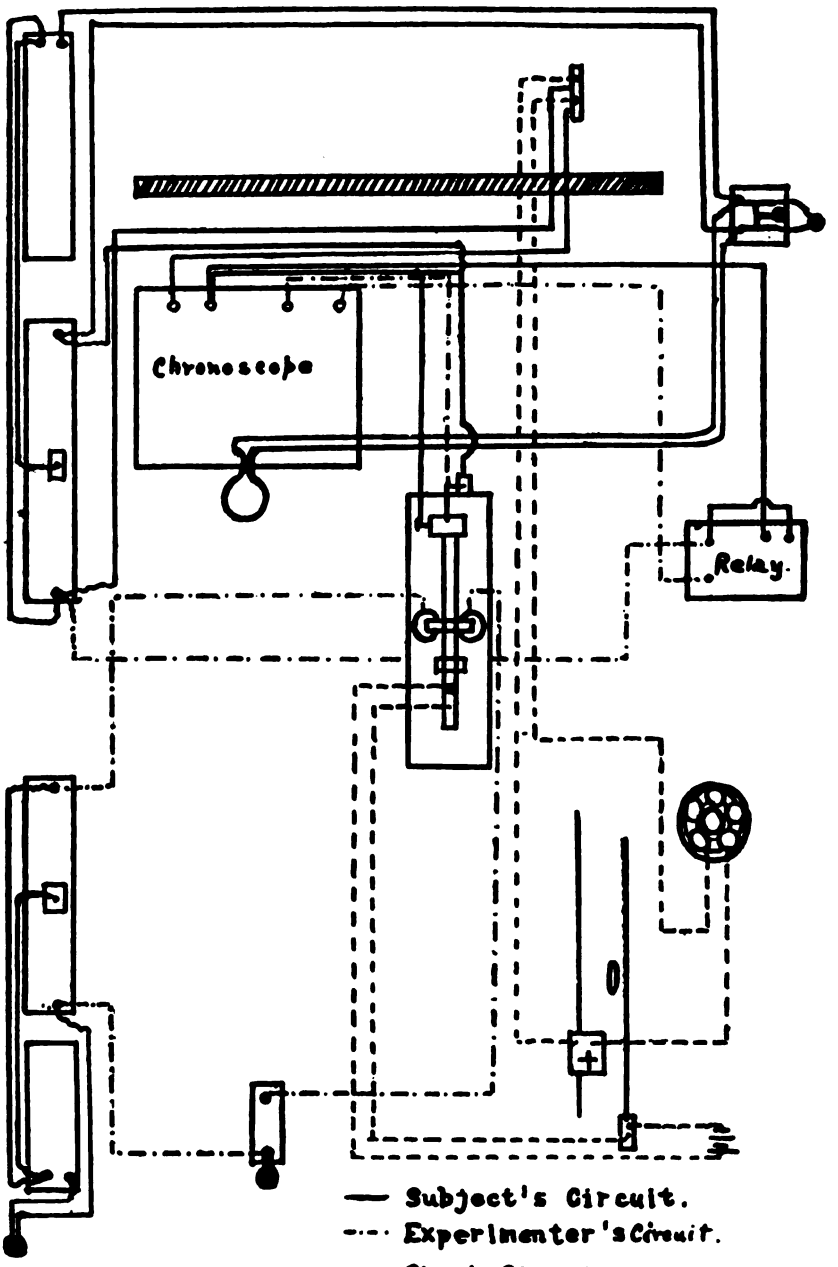


FIG. 1. DIAGRAM OF APPARATUS

its base where the circuit split, one-half leading back to the chronoscope and completing what may be designated as the subject's circuit. The other lead from the mercury cup was taken to the electromagnets of the chronoscope, through a relay and back to the rheostat. By reference to Fig. 1 it will be apparent that by wiring in this manner it was impossible for premature reactions to be recorded unless the sound hammer had actually struck the base and completed the experimenter's circuit, under which condition a premature reaction would affect the reading of the chronoscope. Then to insure a sufficiently long make and break interval in the operator's circuit, a simple relay was included. The circuit from the upper electromagnets of the chronoscope was taken through the relay and back to the rheostat. Thus arranged, the key, which gave the sound stimulus and opened the circuit through the upper magnets of the chronoscope was made to open or close, as conditions warranted, the circuit through the electromagnets of the relay. The armature of the relay when released again made the circuit through the upper magnets. Thus this set of magnets was ready to draw the chronoscope armature up as soon as its release was effected from the lower pair of magnets.

From another source, a current reduced as before to four volts by being passed through a bank of lamps and a rheostat, was taken to the electromagnets of the sound hammer. In this line a simple switch was cut-in which served as the experimenter's control of the stimulus.

Another direct circuit was taken from the University current, passed through a bank of four 40 watt lamps arranged in parallel, to a trip-switch device located on a wooden bar. This switch was connected with the subject's key, wired and insulated so that the current passed through two small copper plates attached to the rubber knob of the telegraph key. By means of this circuit the inclusion of the electric shock was controlled. The trip-switch device used to throw the current on or off acted practically instantaneously.

A three cell battery circuit connected to an electromagnet magnetized it so that it held a pendulum at one end of its arc. In this circuit was placed a mercury cup acting as a make and break or more simply as a switch operating with the throw of the sound hammer arm. The break of the circuit was thus simultaneous with the giving of the stimulus. The small mercury cup was placed directly under the arm of the sound hammer with a wire leading to the electromagnet. The other wire in the same circuit, dipping into the mercury cup, was fastened to the arm of the hammer, but insulated from it. As the end

of the wire was lifted out of the mercury cup with the downward stroke of the hammer, the circuit was broken as the stimulus was given. Breaking the circuit demagnitized the electromagnet, thereby releasing the pendulum which in its swing tripped a small pin in the trip-switch mentioned above to throw the current into the plates on the subject's reaction key. The bar upon which the trip-switch was fastened was fully graduated with a time scale in sigma. The switch could be adjusted upon this scale whenever it was necessary to vary the length of the interval between the giving of the stimulus and the giving of the electric punishment.

The subject was entirely screened off from all the apparatus except his reaction key. There were absolutely no distracting sounds of any sort whereby the subject might be enabled to anticipate the stimulus. The only warning of a stimulus was the ready signal, which will be discussed later. Such sound as was made by the relay, the click of the armature against its magnets, was inaudible to the subject as it occurred simultaneously with the stronger sound stimulus. Any slight noise made by the experimenter in moving or throwing the switches was effectively concealed by the constant hum of the chronoscope mechanism.

By having the apparatus arranged in the manner outlined, the use of the ordinary house current was permitted since any variation in the strength of the current affecting both circuits simultaneously would be of no consideration. Further, the use of storage or gravity batteries, always of considerable inconvenience, was rendered unnecessary and the constant testing with a control chronometer previously necessitated by the effect of slight variations in the current was eliminated. The apparatus in this form was apparently mechanically perfect and did not permit any considerable latency in any part of the arrangement.

It was, however, desirable to know that the chronoscope was recording perfectly both at the beginning and at the end of the experimental period. For this purpose some reliability measures were made. To test the reliability of the chronoscope the Cattell gravity chronometer was employed. This control instrument was set for an interval of two hundred sigma. The heavy weight held at the top of a two meter column by large magnets, when released started and stopped the recording mechanism of the chronoscope by means of two wheel contacts set at the proper angles. Checking the readings of the chronoscope with the standard interval determined the reliability of the chronoscope. The chronometer has been figured and described

by Cattell[3]. He found it to be very reliable and in a test[4] of three successive series of ten trials each, with a normal time of 100 sigma, readings whose mean variations were 0.54, 0.64, and 0.56 sigma were obtained. The reliability measures for the chronoscope at the beginning of the present investigation were the following:

199.7	M.V. .9
198.8	M.V. 1.8
199.7	M.V. 1.4

Each figure represents the average of ten trials. The standard readings were 200 sigma. Taking the mean of these readings gives as the measure of accuracy the figure 199.7 sigma with an M.V. of 1.1. At the conclusion of the experimentation the chronoscope was again tested in the same manner. The results were:

198.1	M.V. .54
198.5	M.V. .50
198.8	M.V. .48

This gives a mean of 198.4 sigma with a mean variation of .54. The conclusion is that the chronoscope was registering perfectly and that discrepancies could not occur from this source.

2. Procedure.

There were essentially three presentations of the sound stimulus, but since the method of presentation was substantially the same in all three it is only necessary to give the fundamental process in detail and note the variations. The normal series of reactions comprised simply the normal reactions of the subject to the sound stimulus without any complicating factors or conditions. The "incentive" series, as they will be designated throughout, comprised those reactions to the sound stimulus after the subject had been informed that the time he took to react would be presented to him before the next succeeding reaction occurred. The time was thus used as an incentive to condition the reactions. The third series, designated as the "punishment" set, involved the following explanation to the subject. "This time you will react as before to the sound stimulus. Use the same movement in releasing the key, only be sure that in holding the key down you have had a finger on each of the two copper plates. You will receive an electric shock only when you do not react quickly enough. That is, if you begin to slow down and take long to react

[3] Catell and Dolley, "On Reaction-times and the Velocity of the Nervous Impulse", Nat. Acad. Sc. Sec. Mem., 7, 397, 1893.

[4] Catell, "Chronoscope and Chronograph", Phil. Stud., 9, 309, 1894.

you will not be able to get away from the key without being shocked, A shock means then that you are not doing as well as you have been doing and that your reaction time is long. Avoid the punishment by speeding up and by maintaining that speed."

The subject, screened off from everything but his reacting key, was placed at one end of the table opposite the experimenter. The experimenter recorded the reading of the chronoscope. As the chronoscope was started a ready signal, simply the word "READY", was given the subject who placed the first two fingers of his right hand upon the knob of the key, pressing it down against its base. The circuit through the electromagnets of the chronoscope was thus closed. After an interval of approximately two seconds, the experimenter closed the control switch which caused the stimulus to be given simultaneously with the closing of the circuit through the upper electromagnets. Completing this latter circuit started the recording mechanism of the chronoscope. The subject, as soon as he perceived the stimulus, released his key thereby breaking the circuit and stopping the recording of the chronoscope. This method of procedure was followed throughout the normal series of reactions.

In the second series of reactions, designated in the statistical tables as the "incentive" set, another step was added to the process used in obtaining the normal reactions. Immediately before giving the signal "Ready", the time of the subject's previous reaction was presented to him by auditory presentation in the form of a statement such as, "Your time was 150 sigma", or later simply as "150", the subject understanding thereby that that was his reaction-time for the trial just completed. Then the signal "Ready", a foreperiod of about two seconds, followed by the sound stimulus to which the reaction was made with the readings recorded as above and the reaction was completed.

With the next series, the punishment set was given, and the procedure became somewhat more complicated for the experimenter. Now the time of the preceding reaction was read and recorded as before, but was not divulged to the subject. The experimenter set the pendulum in place; threw the trip-switch, as shown in the apparatus diagram, into position; gave the ready signal as the chronoscope was started; waited the proper length of time, then gave the stimulus. The pendulum in its swing tripped the switch which opened the punishment circuit through the subject's key. Whether the subject received the shock or not depended solely upon the quickness with which the subject released his key and got away from it. While the time of connecting the punishment circuit with the subject's key

could be varied, the strength of the current was not varied but remained constant throughout the experiment.

3. Standard Conditions.

Standardized conditions, as determined by previous investigators in the field of reaction-time, were adhered to throughout the experiment. It has been determined most recently by Froeberg[5] that the reaction-time increases as the intensity of the stimulus decreases. Consequently the intensity of the sound was kept as constant as possible. Further the nature of the apparatus itself was such that the duration of the stimulus was constant during all periods of work. This factor was of importance, for a variation in the duration of the stimulus has been cited as a cause of varied times of reaction. Referring to Froeberg again, he concludes that the effects of variation of the duration on the reaction-time are "(1) that the time of reaction increases with a decreasing duration of the stimulus, (2) that the time of reaction increases arithmetically as the duration of the stimulus decreases geometrically."

In determining the optimum length of the foreperiod the history of this particular phase of reaction-time was canvassed very thoroughly. The greatest of care was exercised in varying the length of the foreperiod so as to keep it within very narrow limits and yet not have it so regular as to enable the subject to feel the rhythm or to obtain aid in any way in anticipating the stimulus. The constant and close attention of the experimenter was required by this factor in order to avoid this possibility as a source of error. Todd[6] has given a very complete resumé of the findings of the investigations of the foreperiod. He says, "The interval of from one and a half to two seconds has been experimentally determined as the most favorable for obtaining regular reaction-time. Martius found that the reactions with an irregular interval between the signal and reaction stimulus result in a time intermediate between that with a regular signal and with no signal at all. Estel has found that the most favorable interval is two and one-fourth seconds; Wundt places it at two and one-half . . . . . and a recent investigator, Breitweiser, has found that the optimal interval between the signal and stimulus, and the one most often preferred by his subjects was a period of from one to four seconds. Bliss states: 'Experiment has shown that when an interval between warning and stimulus is always the same the mind is soon able to estimate the interval correctly and always reacts just at that

[5] Froeberg, "The Relation between the Magnitude of the Stimulus and the Time of Reaction", Archives of Psych., 8, 1907.

[6] Todd, "Reaction to Multiple Stimuli", Archives of Psych., 25, 1912.

time whether it hears the stimulus or not.' Woodworth: 'If the stimulus follows the signal at an irregular interval, the reaction-time is not so short as when the procedure is regular. If, indeed, the procedure is so regular that the moment of the stimulus can be exactly anticipated, the movement may be made to coincide in time with the stimulus, and the whole character of the experiment be thus changed. This result is avoided by varying the preliminary interval within narrow limits. The most favorable interval between the ready signal and the stimulus is one or two seconds; a shorter time does not allow the subject to prepare himself fully for the stimulus, while a longer period than two seconds allows more time than was needed and so affords a chance for wandering of the attention.'" The more recent tendency seems to be in favor of an interval between the ready signal and the stimulus of from two to four seconds. Supporting this view is the statement from Woodworth and Poffenberger[7] that the investigations have shown that the most favorable interval is usually from two to four seconds, varying somewhat from one individual to another. That is, the reactions were quicker when the interval was of this length than when it was as short as one second or as long as five seconds. The existence of a most favorable interval seems to be established. One second is almost too short a time to enable good preparation, while an interval of over four seconds causes the state of readiness to fluctuate.

Before taking any records in this experiment the subjects were instructed and practiced in the kind of movement to make in releasing the key in response to the stimulus. It was found that all of the subjects felt more comfortable and believed that they could react more quickly and with less variation by simply lifting the fingers from the key instead of drawing away from it. The former involved movement of the fingers and wrist only, whereas the latter caused the whole arm and even a movement of the body to be included. Consequently the subjects were instructed to react by lifting the fingers; to use a sensory type of reaction and to adhere to that mode of response consistently throughout the experiment. Several investigators, Wundt, Titchener, Breitweiser, and others have found that there is an appreciable difference between the two types of reacting. The motor type, it is claimed, has the advantage over the sensory type, but as far as this investigation is concerned this discussion does not pertain since only the one type of reaction was involved and further since comparisons of the several series of reactions was made for each subject and not between subjects.

[7] Woodworth and Poffenberger, "Experimental Psychology" (Mimeog.), 1920, 191.

CHAPTER III.

THE RESULTS.

The results of the experiment are stated in terms of the improvement caused by the factors of incentive and punishment. The time of the reaction is given in sigma. There are three types of reactions, the normal, the incentive, and the punishment. Each one of these types is in turn divided into three series based upon the order of presentation in the daily set of reactions. The order of presentation was rotated in order to avoid a time error as well as to distribute practice and fatigue effects equally throughout all of the series of reactions. Each figure in these tables of results is the average of 350 single reactions for subjects (N. and J.) and the average of 500 single reactions for subject H. In treating the results two groupings are possible either one of which shows the gain or improvement in the time of a reaction by an incentive or by punishment. The first grouping is determined by the time of presentation at each daily sitting. For instance, the first day's results were obtained in the order, normal first, incentive set second, and punishment set third; the second day's results were in the order, incentive set first, punishment set second, and the normal third. Then the third day's results were in the order, punishment set first, normal second, and incentive set third, and so on through the experiment. The second method of grouping is obtained by taking all the several types of reactions given first throughout the experiment in one group, those given second in another group, and those given third in a final group. Then comparing the three types of reactions within each classification, the amount of improvement is obtained for each position.

The general results of the experiment are given in the following tables. Table I is a summary table of the original measures given completely in another connection later, with the additional data from which the reliability measures are calculated. An inspection of the table will show the surprisingly large saving in the time of the response to the sound stimulus caused by the accompanying factors, incentive and punishment. The conclusion that may be formulated from this table is that the introduction of a positive incentive is a powerful factor in reducing a normal reaction-time, which ordinarily is supposed to be as fast as the subject is able to react. That is, when measuring the

reaction-time of subjects who are coöperating and seriously doing their best, a normal reaction is secured representing their practiced rate of responding to sound. But here the introduction of an incentive reduces the time below this normal and speeds up a regular response already thought to be at its best. The average was found to be six percent of the normal time.

Then introducing a negative incentive, that is, something to be avoided, in this case the electric shock, the previous low normal is reduced fifteen percent. In other words, a positive incentive secures a reduction of 8.8 sigma and a negative incentive 21.4 sigma.

Explanation of Table I: In this table the records of the reactions are given in the three orders in which they were presented to the subjects. Order I corresponds in sequence to those obtained on the first day; Order II to those of the second day, and Order III to those of the third day as explained above. The time is given in sigma. Each figure represents the average of 350 reactions for subjects N. and J. For subject H. the average is based upon 500 reactions. The general or total average of each type of reaction is given at the bottom of each column. The M.V.'s are the average deviations of the reactions grouped by fifties from the average. The P.E.'s are the probable errors of the averages, derived by the formula,

$$\text{P.E.} = \frac{.8453\ \text{M.V.}}{\sqrt{n}}$$

TABLE I.

Normal, Incentive, and Punishment Reaction Series.

ORDER I.

	Normal		Incentive		Punishment	
	Av.	M.V.	Av.	M.V.	Av.	M.V.
N.	144.1	2.8	135.8	2.2	125.4	1.9
		P.E. .89		P.E. .70		P.E. .60
H.	144.1	1.5	134.9	2.0	118.0	1.5
		P.E. .40		P.E. .53		P.E. .40
J.	141.0	4.8	133.6	3.3	123.8	1.6
		P.E. 1.53		1.05		.51

ORDER II.

N.	144.5	4.3	136.2	1.8	124.1	2.4
		P.E. 1.37		.57		.76
H.	147.4	4.5	134.8	2.4	119.3	2.1
		P.E. 1.2		.64		.56
J.	141.7	1.8	134.1	2.5	123.6	1.3
		P.E. .57		.79		.41

ORDER III.

N.	144.9	4.1	135.8	2.2	125.5	1.9
		P.E. 1.31		.70		.60
H.	146.0	2.2	135.9	2.6	118.6	2.3
		P.E. .59		.69		.61
J.	142.0	3.7	134.7	3.3	123.7	1.9
		P.E. 1.18		1.05		.60
Average	143.9		135.0		122.3	
		P.E. .45		P.E. .20		P.E. .67

7 %

From the averages given in the previous table, the amount of improvement both in terms of sigma and in terms of percent has been calculated. Table II gives this improvement for each subject in terms of sigma. The three orders of presentation are given.

TABLE II.

		Improvement by Incentive	Improvement by Punishment
(H)	*Normal, Incentive, Punishment Order.*		
Normal	144.1		
Incentive	134.9		
Punishment	118.0	9.2	26.1
(N)			
Normal	144.1		
Incentive	135.8		
Punishment	125.4	8.3	18.7
(J)			
Normal	141.0		
Incentive	133.6		
Punishment	123.8	7.4	17.2
	Incentive, Punishment, Normal Order.		
		Improvement by Incentive	Improvement by Punishment
(H)			
Normal	147.4		
Incentive	134.8		
Punishment	119.3	12.6	28.1
(N)			
Normal	144.5		
Incentive	136.2		
Punishment	124.1	8.3	20.4
(J)			
Normal	141.7		
Incentive	134.1		
Punishment	123.6	7.6	18.1
	Punishment, Normal, Incentive Order.		
		Improvement by Incentive	Improvement by Punishment
(H)			
Normal	146.0		
Incentive	135.9		
Punishment	118.6	10.1	27.4
(N)			
Normal	144.9		
Incentive	135.8		
Punishment	125.5	9.1	19.4
(J)			
Normal	142.0		
Incentive	134.7		
Punishment	123.7	7.3	18.3

Interpreting the results of the foregoing table in terms of percent of improvement and grouping in the same general way, the figures are as shown in the next table.

TABLE III.

Improvement by Incentive and Punishment in Percent.

When the series was presented in the Normal, Incentive, Punishment order the percent improvement was:

Subject	By Incentive	By Punishment
H.	6.0	18.1
N.	5.7	12.9
J.	5.2	12.1

When the series was presented in the Incentive, Punishment, Normal order the percent improvement was:

Subject	By Incentive	By Punishment
H.	8.5	19.0
N.	5.7	14.1
J.	5.2	12.7

When the series was presented in the Punishment, Normal, Incentive order the percent of improvement was:

Subject	By Incentive	By Punishment
H.	6.9	18.7
N.	6.2	13.3
J.	5.1	12.8

Grouping together the reactions according to the second method of grouping mentioned previously (p. 25), slightly varied percents of improvement are obtained. Whenever the groups were presented first in the daily series the percents were:

Subject	By Incentive	By Punishment
H.	6.1	17.6
N.	5.4	12.9
J.	4.8	12.2

Grouping as above all the sets presented second throughout, the percents were:

Subject	By Incentive	By Punishment
H.	7.6	18.2
N.	6.8	14.3
J.	5.9	12.9

Grouping the sets always presented third in the series, the percents were:

Subject	By Incentive	By Punishment
H.	7.8	19.9
N.	6.0	13.2
J.	4.9	12.6

Figures 2, 3 and 4, are graphic representations of the percent of improvement as given in the first part of Table III (p. 29).

To further summarize the results and show more concisely the influence that incentive and punishment may have upon one's normal reaction-time, the percent of improvement brought about by the conditions of the experiment, for all subjects and under all conditions were grouped together, giving the following:

TABLE IV.

	Percent improvement by Incentive	Percent improvement by Punishment
Order No. 1	5.4	14.2
Order No. 2	6.7	15.1
Order No. 3	6.2	15.2

In this table "Order No. 1", means that all of the three types of conditioning the reaction were presented first in some series throughout the experiment and that all these were grouped together. The other two orders were derived in a similar manner.

Then summarizing the first section of Table III the following results in which "Order N." means that the series were presented in the order Normal, Incentive, and Punishment. "Order I." means that the Incentive set was given first, the Punishment set second, and the Normal set third. "Order P." means that the Punishment set was first, the Normal second, and the Incentive set third. The groupings are as above and obtained in the same manner.

	Percent improvement by Incentive	Percent improvement by Punishment
Order N.	5.6	14.3
Order I.	6.5	15.2
Order P.	6.0	14.9

Up to this point comparisons have been made on the basis of the position of the reaction sets in the series and remarkably large gains have been determined. Another method of showing approximately the same substantial improvement may be applied by basing the comparisons upon the total number of reactions obtained for each subject. These results have been worked out both in terms of sigma and in terms of percent of improvement. Table V gives a comparison in sigma of the amount improved as determined from the total number of reactions of each of the three subjects. For subject H. the reactions total 4500, while for subjects J. and N. the reactions total 3150 each. In this table the comparisons are given for each type regardless of the position in the daily series.

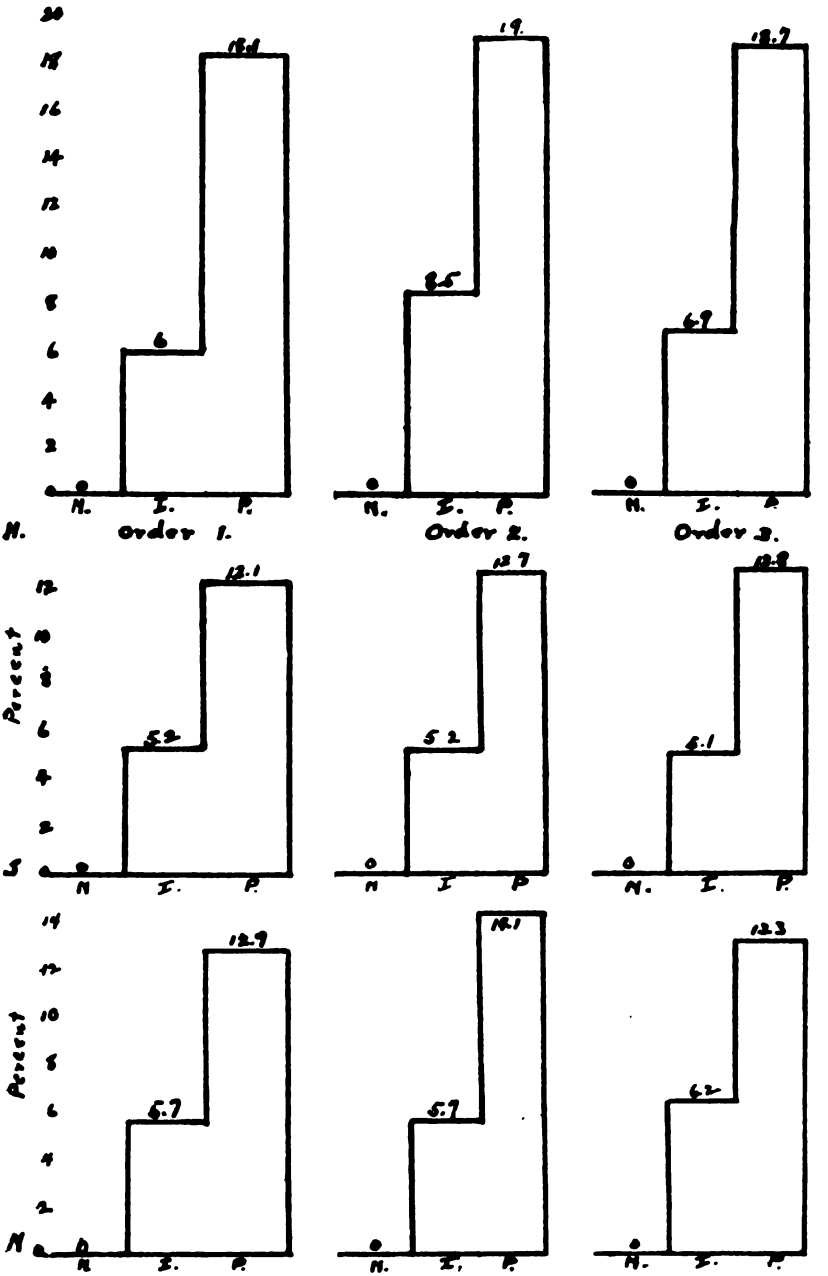


Fig. 2,3,4. Percent Distribution of Improvement.

TABLE V.

		Improvement by Incentive	Improvement by Punishment
Subject H.			
Normal Average	145.83		
Incentive "	135.20		
Punishment "	118.63	10.63	27.20
Subject N.			
Normal Average	144.50		
Incentive "	135.93		
Punishment "	125.00	8.57	19.50
Subject J.			
Normal Average	141.23		
Incentive "	134.13		
Punishment "	123.70	7.10	17.53

Converting these results into percents of improvement, the following is obtained. In this table the total average improvement of all the subjects, regardless of the position in the series in which the conditioning factors were given, may be thus stated.

TABLE VI.

Improvement by Incentive	6.0%
Improvement by Punishment	14.8%
Stating the same in terms of sigma:	
Improvement by Incentive	8.8
Improvement by Punishment	21.4

If practice has a dominant influence upon the time of reaction, then the reactions at the end of the day's work would show correspondingly better time than at the beginning of the work period. Similarly, the reactions at the end of the complete series of reactions would be quicker than those at the beginning. Bearing practice in mind one would expect to find greater improvement when any one type of reaction is in the last position in the series than when it is in any of the preceding positions. That is, the punishment set, for instance, when given last in the series ought to show greater improvement over the normal set than the punishment set does when given first. But what do the figures show? Referring to Table II, in the punishment type the improvement over the normal is 27.4 sigma when the punishment set is given first in the series, but only 26.1 sigma when given last. Likewise punishment in the second or intermediate position substantiates this conclusion. The improvement when given second in the series is 28.1 sigma and 26.1 sigma when given last.

A similar comparison of the amount of improvement may be drawn between the punishment set and the incentive type of reaction with similar results, verifying the premise that practice was entirely eliminated before obtaining any results and that any improvement noted was not due to the influence of practice. A survey of the figures of the other subjects show the same general result. With the amount of improvement as affected by position in the series as the criterion, it may justly be concluded that practice was entirely eliminated. The results, therefore, are determined from reactions obtained at the apparent physiological limit.

Figure 5 is a graphic representation of a typical case showing that practice was entirely eliminated. The three types of reaction proceed each at a distinct level, but parallel to each other. There is no trace of the customary sharp decline apparent in the usual practice curve. Each point on the curve is the average of ten reactions.

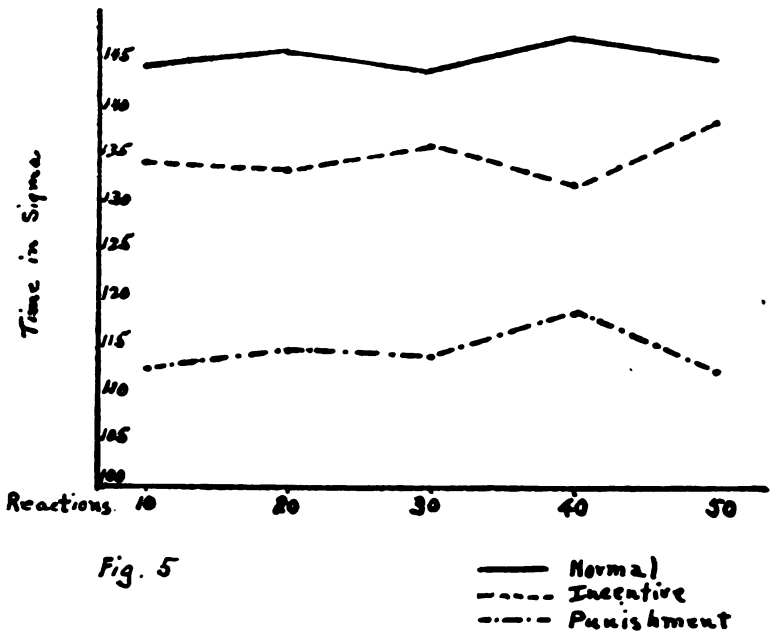


Fig. 5

1. Interpretation of Results in Terms of Reliability.

The consideration of the improvement factors of the experiment may be terminated at this point and the attention directed instead to the determination of the reliability of the measures obtained. Thus far the average time has been given and compared with other averages of the various series, but nothing in the way of reliability of these averages has been stated.

Turning to the original measures themselves, as shown in Tables VII, VIII, and IX, it is to be noticed that each figure represents the average of fifty reactions. The mean variations are the average deviations of the fifty single or individual reactions from the average. The mean variations are quite low in most cases, showing that the subjects were well trained and had reached the practice level before any reactions were recorded. It is usual to measure the reliability of an average by the size of the probable error. The P.E.'s in these records are the probable errors of the averages found according to the formula:

$$\text{P.E.} = \frac{.8453 \quad \text{M.V.}}{\sqrt{n}}$$

in which the number of cases, "n", equals 50.

In Fig. 6 is shown the distribution of the single reactions of the three types, the normal, the incentive, and the punishment. The reactions were grouped according to the limits within which they fell and for purposes of distribution these groups were of five sigma differences. The total number of single reactions included in each of the curves of the distribution is 3600. The curves show very clearly that the three types of reaction were on distinct levels, though with some overlapping. The median for the normal is 144, for the incentive is 135, and for the punishment is 122. It is also noticeable that in both the normal and the incentive there is more spread than in the punishment set. In the latter the cases are grouped more which gives the curve the higher and narrower shape. The range in this case is 40, whereas it is 60 in the incentive and 65 in the normal.

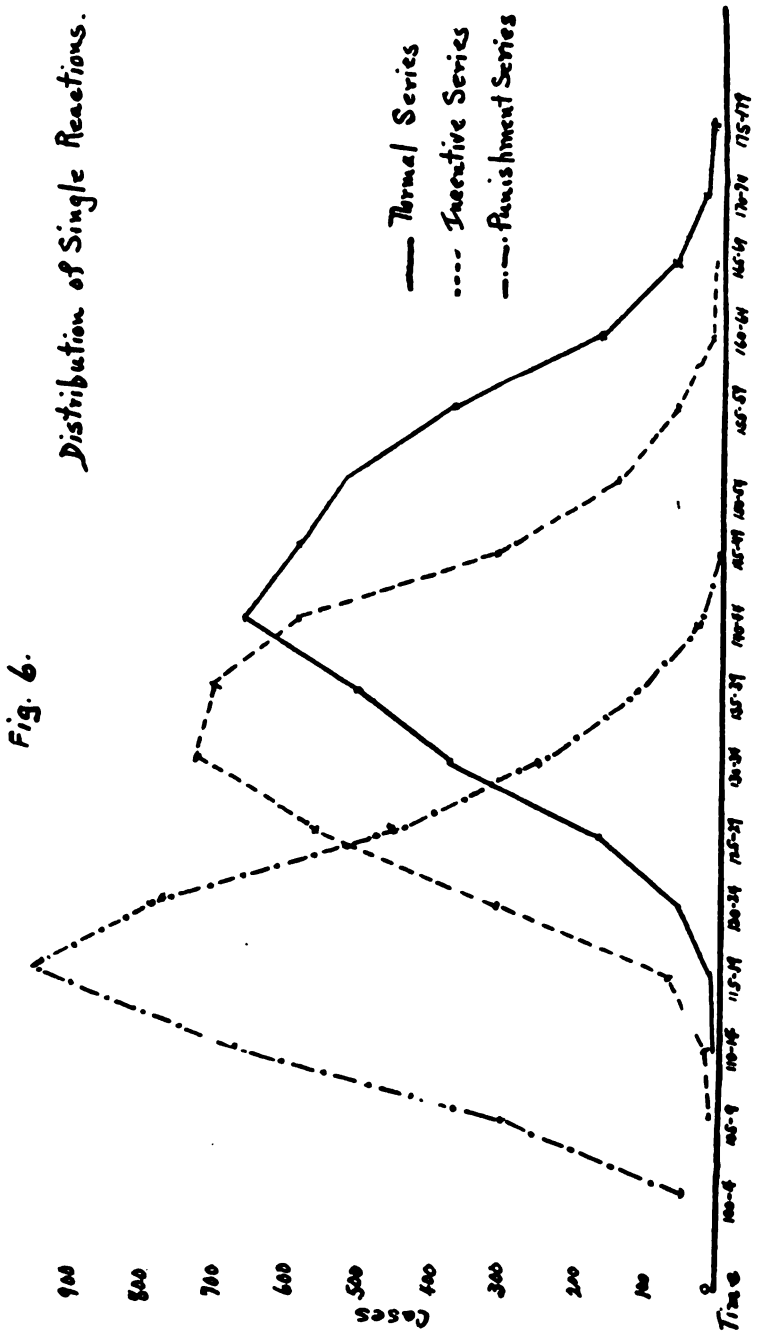


Fig. 6.
Distribution of Single Reactions.
Normal Series
Incentive Series
Punishment Series
Cases
Time

TABLE VII.

Each number in the table is the average time of fifty reactions. The number in a parenthesis designates the position in each daily series of reactions in which the normal, incentive, or punishment set was given. (Subject H.)

Normal		Incentive		Punishment	
(1)	Av. 144.2	(2)	133.0	(3)	120.3
	M.V. 7.5		5.5		7.2
	P.E. .89		.65		.85
(3)	148.1	(1)	133.0	(2)	121.0
	8.1		6.7		5.2
	.96		.79		.62
(2)	145.6	(3)	131.8	(1)	121.5
	6.8		8.9		6.5
	.81		1.06		.77
(1)	138.1	(2)	132.1	(3)	117.4
	8.3		6.1		4.7
	.99		.72		.56
(3)	152.3	(1)	141.1	(2)	118.3
	7.2		6.1		5.8
	.85		.72		.69
(2)	141.7	(3)	133.4	(1)	113.2
	8.7		8.8		6.6
	1.03		1.05		.78
(1)	144.0	(2)	137.9	(3)	119.2
	8.5		8.5		9.1
	1.01		1.01		1.08
(3)	158.5	(1)	133.4	(2)	123.4
	7.4		7.9		7.8
	.88		9.4		.93
(2)	145.4	(3)	131.9	(1)	114.2
	7.5		5.6		5.7
	.89		.66		.68
(1)	145.0	(2)	134.6	(3)	114.3
	7.8		5.7		4.5
	.93		.68		.53
(3)	153.2	(1)	134.8	(2)	124.3
	7.4		6.2		6.3
	.88		.74		.75
(2)	150.3	(3)	137.6	(1)	118.8
	5.7		8.1		5.8
	.68		.96		.69
(1)	142.7	(2)	133.1	(3)	117.6
	8.7		5.1		5.6
	1.03		.60		.66

Normal		Incentive		Punishment	
(3)	142.4	(1)	130.3	(2)	116.2
	12.3		8.1		5.5
	1.47		.96		.65
(2)	144.0	(3)	138.4	(1)	118.2
	6.0		8.1		6.5
	.71		.96		.77
(1)	144.1	(2)	134.9	(3)	120.7
	6.8		5.4		5.4
	.81		.64		.64
(3)	145.2	(1)	137.0	(2)	115.6
	7.2		7.2		4.4
	.85		.85		.52
(2)	148.0	(3)	136.0	(1)	117.3
	8.2		5.4		4.5
	.97		.64		.53
(1)	145.5	(2)	136.1	(3)	118.7
	6.9		5.4		5.1
	.82		.64		.60
(3)	144.0	(1)	136.3	(2)	118.0
	7.7		6.1		4.4
	.91		.72		.52
(2)	149.7	(3)	141.5	(1)	119.5
	6.8		5.7		5.7
	.81		.68		.68
(1)	145.1	(2)	138.3	(3)	118.0
	6.9		5.7		5.0
	.82		.68		.59
(3)	147.5	(1)	137.4	(2)	118.8
	8.2		6.0		4.5
	.97		.71		.53
(2)	146.9	(3)	139.5	(1)	123.6
	7.8		7.0		5.1
	.93		.83		.60
(1)	145.6	(2)	131.7	(3)	119.1
	5.8		6.0		5.0
	.69		.71		.59
(3)	142.0	(1)	131.1	(2)	119.2
	5.8		5.2		5.0
	.69		.62		.59
(2)	146.2	(3)	134.4	(1)	120.4
	7.0		6.4		5.2
	.83		.76		.62
(1)	147.6	(2)	138.1	(3)	115.0
	6.6		7.7		7.0
	.78		.91		.83
(3)	140.9	(1)	134.4	(2)	118.3
	5.6		6.2		5.3
	.66		.74		.63
(2)	142.3	(3)	135.0	(1)	120.0
	7.0		6.7		4.5
	.83		.79		.53

TABLE VIII.

Each figure in this table is the average of fifty reactions. The number in a parenthesis designates the position in each daily series in which the normal, the incentive, and the punishment set of reactions to the sound stimulus was presented. (Subject N.)

Normal		Incentive		Punishment	
(1)	Av. 141.4	(2)	133.3	(3)	123.3
	M.V. 6.7		6.7		8.4
	P.E. .79		.79		1.00
(3)	139.0	(1)	135.5	(2)	123.6
	8.1		6.3		6.1
	.96		.75		.72
(2)	137.3	(3)	132.5	(1)	121.9
	6.4		6.0		6.2
	.76		.71		.74
(1)	140.0	(2)	135.5	(3)	126.1
	6.6		7.7		6.4
	.78		.91		.76
(3)	150.0	(1)	140.6	(2)	127.4
	7.0		7.9		6.4
	.83		.94		.76
(2)	144.6	(3)	133.0	(1)	125.1
	6.1		6.8		5.3
	.72		.81		.63
(1)	143.8	(2)	138.2	(3)	130.2
	6.6		6.9		6.0
	.78		.82		.71
(3)	144.3	(1)	137.1	(2)	127.8
	6.6		7.6		6.6
	.78		.90		.78
(2)	145.5	(3)	136.0	(1)	123.0
	7.7		7.0		6.3
	.91		.83		.75
(1)	149.8	(2)	141.4	(3)	126.8
	8.1		7.6		6.4
	.96		.90		.76
(3)	146.3	(1)	137.2	(2)	123.4
	5.7		6.4		6.1
	.68		.76		.72
(2)	151.0	(3)	141.5	(1)	130.8
	6.0		6.9		6.0
	.71		.82		.71
(1)	146.6	(2)	133.6	(3)	124.1
	5.3		6.5		5.1
	.63		.77		.60

Normal		Incentive		Punishment	
(3)	135.4	(1)	131.5	(2)	120.0
	7.8		7.0		6.7
	.93		.83		.74
(2)	139.2	(3)	132.6	(1)	125.4
	6.3		6.6		6.2
	.75		.78		.74
(1)	141.4	(2)	135.3	(3)	123.3
	7.5		7.4		5.0
	.89		.88		.59
(3)	146.9	(1)	136.6	(2)	120.7
	8.2		7.7		5.6
	.97		.91		.66
(2)	148.8	(3)	136.0	(1)	126.0
	6.0		7.6		6.2
	.71		.90		.74
(1)	145.9	(2)	133.7	(3)	124.2
	8.6		8.0		5.9
	1.02		.95		.70
(3)	150.2	(1)	135.5	(2)	125.8
	7.1		5.5		4.6
	.84		.65		.59
(2)	148.4	(3)	130.2	(1)	126.6
	6.4		6.3		6.3
	.76		.75		.75

TABLE IX.

Each number is the average in sigma of fifty reactions. The number in a parenthesis designates the position in each daily series in which the normal, the incentive, and the punishment set of reactions to the sound stimulus was presented. (Subject J.)

Normal		Incentive		Punishment	
(1)	Av. 137.3	(2)	133.5	(3)	123.6
	M.V. 6.2		7.2		5.9
	P.E. .74		.85		.70
(3)	137.0	(1)	132.3	(2)	124.6
	7.6		6.7		5.2
	.90		.79		.62
(2)	139.0	(3)	131.6	(1)	124.2
	6.9		7.0		6.0
	.82		.83		.71
(1)	140.2	(2)	132.3	(3)	123.6
	9.4		6.0		5.8
	1.12		.71		.69
(3)	143.1	(1)	132.9	(2)	123.1
	8.9		6.1		6.7
	1.06		.72		.79

Normal		Incentive		Punishment	
(2)	136.7	(3)	129.4	(1)	120.0
	6.5		8.6		5.3
	.77		1.02		.63
(1)	139.6	(2)	132.4	(3)	125.2
	7.3		7.2		6.9
	.87		.85		.62
(3)	142.8	(1)	136.9	(2)	123.5
	8.6		7.5		7.0
	1.02		.89		.83
(2)	143.3	(3)	133.4	(1)	121.8
	8.9		6.7		5.4
	1.06		.79		.64
(1)	136.1	(2)	127.6	(3)	118.7
	8.4		6.3		5.5
	1.00		.75		.65
(3)	140.8	(1)	129.7	(2)	120.3
	8.1		7.5		5.3
	.96		.89		.65
(2)	147.0	(3)	135.8	(1)	126.3
	9.3		7.1		6.3
	1.11		.84		.75
(1)	157.0	(2)	145.0	(3)	127.0
	7.2		9.3		7.7
	.85		1.11		.91
(3)	141.1	(1)	132.6	(2)	125.6
	7.8		6.4		5.2
	.93		.76		.62
(2)	138.7	(3)	132.9	(1)	122.8
	9.4		6.2		6.7
	1.12		.74		.79
(1)	142.0	(2)	134.1	(3)	124.0
	8.7		7.6		6.2
	1.03		.90		.82
(3)	141.6	(1)	134.6	(2)	123.1
	7.3		7.5		5.7
	.87		.89		.68
(2)	148.8	(3)	144.9	(1)	123.6
	7.8		6.0		6.9
	.93		.71		.82
(1)	135.2	(2)	130.9	(3)	125.0
	8.3		5.9		6.7
	.99		.70		.79
(3)	146.0	(1)	140.0	(2)	125.5
	6.1		6.9		6.5
	.72		.82		.77
(2)	140.7	(3)	135.3	(1)	127.4
	8.3		8.6		8.0
	.99		1.02		.95

In determining the reliability of the sigma improvement resulting from the conditioning factors of incentive and punishment as noted in the preceding tables, the P.E.'s of the differences were calculated. According to Thorndike[1], "The unreliability of a difference between A and B equals the square root of the sum of the square of the unreliability of A and the square of the unreliability of B." The probable error of the difference was calculated from the formula given in the following "Reliability Table." The M.V.'s used in this process are not the variations of the individual measures from the average, but the variation of the average measure of 350 and 500, according to the subject, reactions from the average of the total group including seven and ten such group averages, depending upon the subject.

Drawing conclusions from the results given in Table X, it can be said that for subject N., the lowest reliability measure gives an actual difference which is 5.5 times the P.E. Interpreting this in terms of probability, there is only one chance in five thousand that the obtained difference may be due to something other than the conditioning factors used in the series. All the other differences for this same subject range entirely beyond any possibility of chance happening.

Subject J.'s differences rank somewhat more within the realm of chance. A P.E. of 1.8 is derived with a difference which is 4.1 times the P.E. But as this is the lowest, it is fair to assume that the difference is real and not invalidated by the possibility of chance.

The results of subject "H" seem entirely beyond any question since the nearest difference is 9.6 times the P.E.

If instead of forming conclusions from the three types of reactions divided into groups according to the position in the series in which each type was presented to the subject, the total number of each kind of reaction serve as the basis for the conclusion somewhat larger probable errors are obtained with a corresponding decrease in the certainty of the differences.

With subject "N" an improvement of 8.5 sigma was developed by incentive, that is, a normal reaction of 144.5 sigma was reduced to 135.9 sigma with a P.E., of the difference of 1.7. This means that the obtained difference was just five times the probable error. Introducing punishment caused a further decrease of 19.5 sigma, or from 144.5 to 125.9 sigma with a P.E. of 1.6. The obtained difference in this case is 12.1 times the P.E.

A normal reaction-time of 141.2 sigma for subject "J" was improved by the incentive 7.1 sigma, bringing the average incentive-

[1] Thorndike, "Mental and Social Measurements", 193.

reaction down to 134.1 sigma. The P.E. of the difference was 1.9, making the actual difference 3.7 times the P.E. This was the lowest probability of any measure in the entire experiment. For the same subject the improvement secured by punishment amounted to 17.5 sigma or from 141.2 to 123.7 sigma, with a P.E. of 1.8 being 9.7 times the P.E. for the difference.

The results of subject H. were reduced from 145.8 to 135.2 sigma by incentive, a difference of 10.6 sigma with a P.E. of 1.5. Here the difference is seven times the probable error. Punishment established the low level in reaction-time in this experiment with this subject. The decrease was from 145.8 to 118.6 sigma, a saving of 27.2 sigma and a P.E. of 1.4. The difference is here 19.4 times the P.E. In the formula used for these results "n" was 21 for the first two subjects and 30 for the last subject.

TABLE X.

Reliability Table.

(Subject H.)					
Normal		Incentive		Punishment	
Average (1)	144.1	(2)	134.9	(3)	118.0
M.V.	1.5		2.0		1.5
Difference			9.2		26.1
P. E.			.67		.56
Average (3)	147.4	(1)	134.8	(2)	119.3
M.V.	4.5		2.4		2.1
Difference			12.6		28.1
P. E.			1.3		1.3
Average (2)	146.0	(3)	135.9	(1)	118.6
M.V.	2.2		2.6		2.3
Difference			10.1		27.4
P.E.			.91		.84
(Subject N.)					
Average (1)	144.1	(2)	135.8	(3)	125.4
M.V.	2.8		2.2		1.9
Difference			8.3		18.7
P. E.			.74		.65
Average (3)	144.5	(1)	136.2	(2)	124.1
M. V.	4.3		1.8		2.5
Difference			8.3		20.4
P.E.			1.5		1.6
Average (2)	144.9	(3)	135.8	(1)	125.5
M.V.	3.9		2.7		1.9
Difference			9.1		19.4
P. E.			1.5		1.3
(Subject J.)					
Average (1)	141.0	(2)	133.6	(3)	123.8
M.V.	4.8		3.3		1.6
Difference			7.4		17.2
P.E.			1.8		1.6
Average (3)	141.7	(1)	134.1	(2)	123.6
M.V.	1.8		2.5		1.3
Difference			7.6		18.1
P.E.			.70		.90
Average (2)	142.0	(3)	134.7	(1)	123.7
M.V.	3.7		3.3		1.9
Difference			7.3		18.3
P.E.			1.2		1.6

In the foregoing table is given the average results for each subject in each series when presented in the three orders. For two of the subjects, N. and J., the average is derived from 350 reactions, while for the third subject H., the average is obtained from 500 reactions. This gives measures in groups numbering 7, 7, and 10, respectively

for the subjects. The M.V. is based on these group deviations from the average. The difference column in the table denotes the actual amount of improvement in sigma due to the conditioning factors in the several types. The difference is in all cases between the normal and incentive in the first column, and between the normal and punishment in the last column. The P.E. of the difference is derived from the formula,

$$\text{P.E.} = \sqrt{\frac{(.8453\ \text{M.V.})^2\text{A.}}{\sqrt{n}} + \frac{(.8453\ \text{M.V.})^2\text{B.}}{\sqrt{n}}}$$

in which "n" is 7 or 10 depending upon the subject.

The number in the parenthesis designates the position in the series in which each type of reaction condition was used, i.e. (1) means that the normal reaction set was taken first, (2) second, and so on.

An interesting point may be worked out in connection with the punishment series of reactions by noting the effect of a shock upon the immediately following reactions. Averaging all the reactions which are recorded just prior to the infliction of a shock, the result is for each of the three subjects 125.5 sigma, 122.2 sigma, and 124.3 sigma respectively. The average of the reaction for which the shock was received was 138.6, 140.3, 140.6 sigma respectively. Taking the average of the first reaction following the shocked reaction the figures are 121.4, 113.2, 123.2 sigma; for the second reaction following the results are, 120.6, 115.1, 120.4. In the case of the first subject it is to be noted that the first reaction immediately following the shock is slightly longer than the second one following. The third subject's results show the same thing. It would seem then for these two subjects that the full benefit of the shock, or rather the full effect of it is not apparent immediately, but is attained in the second reaction following the shock. The reactions recorded third after the punishment average 120.7, 118.3, and 120.2 sigma. From then on there is a gradual slowing up until the next reaction receiving a shock speeds the process again. Then this process of reduction is repeated and holds for a time. Judging from this particular case it would seem that the reactions for which a shock is given become slower as the series advances, i.e., the last shocked reaction is longer than the first one. But this is not the general case. The length of the shock reactions vary throughout the set. There does not appear to be an undue slowing as a reaction following a shock. On the contrary, while there is a slowing it is very gradual and is about the same as the slowing appearing in the reactions taken before any shock is given.

Fig. 7.

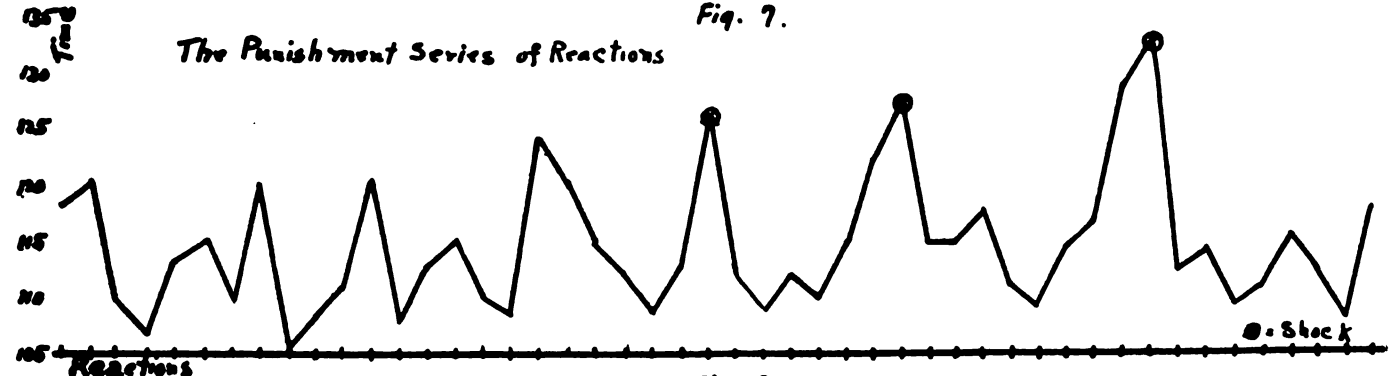
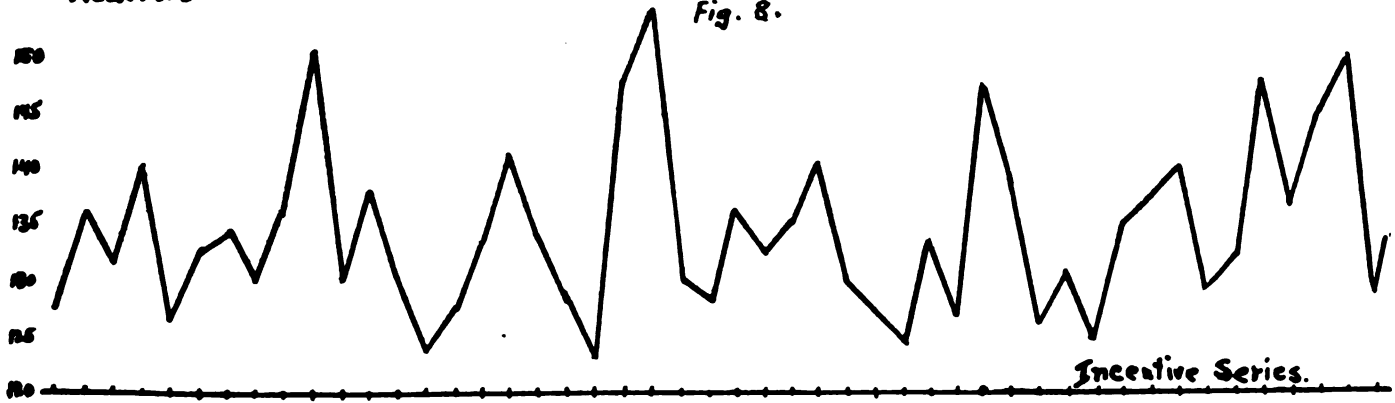

The Punishment Series of Reactions
130
125
120
115
110
105
Reactions
= Shock

Fig. 8.


150
145
140
135
130
125
120
Incentive Series.

The graph in Fig. 7 is a typical case showing the effect of the shock. It was selected at random from all the results and is a fair sampling of a group. The graph also shows that while the shock reduces the time of a reaction very rapidly, the speeding up effect of the punishment does not hold very long. A punishment series taken one day creates a low level in the subject's reacting for that day, but this low level does not hold over to the next day. The reduction is not permanent. In the same way the duration of the effect of a slow reaction in the incentive series is very brief. Summarizing the data given and putting it in tabular form we have:

Subject.	Last before Shock	Shock	1st fol. Shock	2nd fol. Shock	3rd fol. Shock
J.	125.5	138.6	121.4	120.6	120.7
H.	122.2	140.3	113.2	115.1	118.3
N.	124.3	140.6	123.2	120.4	120.2

A similar procedure was followed with regard to the incentive series. The first point to be answered was this. Does an exceedingly slow reaction tend to discourage or inhibit the subject so that the following reactions are slower, or does it give him a greater drive and stimulate him to speed up and make quicker reactions? The records worked over as above in the punishment series give these results. The average of the reactions immediately preceding a very slow reaction is for each subject, 135.3 sigma, 139.7 sigma, and 137 sigma. Then averaging the reactions considered as slow the result is, 149.3 sigma, 156.4 sigma, and 155 sigma respectively. The first following reaction averages 133.1, 137.4, and 132.8 sigma. The second following reaction averages 132.1, 132.2, and 134.8 sigma. The third reaction following averages 134.2 sigma, 135.3 sigma, and 136.3 sigma.

Here again the conclusion may be formulated that the incentive that the reaction was slow acts as a facilitation in the same manner as the shock does. The maximum effect is not immediate, but rather delayed, since the second following reaction is shorter than the first following. However, this is not conclusively established as the third subject's results show that the full facilitation effect is reached immediately. There is then in general a stimulation brought about by the incentive that the reaction was slow, but with variation in the time within which the incentive attains its maximum strength. Fig. 8 is a graphic representation of the incentive series showing the rapid decrease after a slow reaction. The high points are the very slow reactions and in each case it will be noticed that the next succeeding re-

actions continue for a time on a new low level. The second high point in the graph shows how the full effect of the slow reaction was not attained until the second reaction following. Grouping the figures in order to make comparison the easier we have:

Subject	Last before Incentive	Slow Reaction	1st fol.	2nd fol.	3rd fol.
N.	135.3	149.3	133.1	132.1	134.2
J.	139.7	156.4	137.4	132.2	135.3
H.	137.0	155.0	132.8	134.8	136.3

A second consideration of the incentive series was whether or not the improvement noted was really due to the incentive. Might not the improvement only be the result of the conversation involved in stating the time to the subject? A few words, of course, but possibly they were enough to keep the subject awake and more attentive than in the normal series where no word was spoken. If that were true then the improvement ought to be the same throughout the series and not vary according as the reaction was slow or fast. The previous paragraph has shown that the reactions following a very slow reaction are very rapid and Fig. 8 shows that there are varying degrees of improvement depending upon the length of the incentive reaction immediately preceding. Since then such differences appear within the series, the improvement must be caused by the motive set up by the incentive and not by the conversation.

The effect of the fast reaction in the incentive series was a slight slowing up of the reaction. The average of the reactions taken just before the fast reactions was 123 sigma, the fast reaction averaged 119 sigma and the average of the reactions immediately following was 132 sigma. Another subject averaged 135 sigma, 120 sigma, and 137 sigma respectively. The third subject gave results in which the average of those before and after a fast reaction were equal, i.e., 130, 119, 130.

Treating the middle range of the incentive series in the same way the results show that reactions which were intermediate between the very fast and the very slow reactions had an alternating facilitating and inhibitory effect. Sometimes the immediately following reactions were slower than the preceding intermediate reactions and again they were faster. It is, therefore, impossible to formulate any more definite conclusion for the intermediate reactions. However, since the effects did vary according as the reactions were fast or slow in this incentive series, it can be said that the incentive used was a real motive affecting the results and not simply conversation.

A comparable effect may be attributed to premature reactions. They do not affect the three types of reactions in the same way for the incentive set shows that the after-effect of a premature reaction is towards an increase in the time of the next reaction rather than to decrease it. But in the other two series, the normal and the punishment, the effect of the premature reaction is to reduce the time of the following responses. On the whole, the premature reactions indicate a facilitating effect similar to the effects of the slow reactions in the incentive set and the punished reactions. The figures for the premature reactions in the three types, normal, incentive, and punishment, are the following:

	Reaction before	Premature	Reaction following
	151.5	67	145.5
Normal	147.0	69	141.1
	152.5	79	144.7
	144.2	82	145.5
Incentive	131.4	66	135.5
	140.0	65	115.5
	134.5	67	132.1
Punishment	117.5	70	115.5
	134.3	75	123.0

These results are the averages of all the premature reactions occurring in the various sets. The subjects are in the order J., H., and N. in each case.

Should the criticism be made that the improvement caused by the incentive and by the punishment might be due to the presence of premature reactions in the results rather than to the influence of the conditioning factors of the experiment, the defence can be made by an appeal to the distribution curve given in Fig. 6. These three curves of the distribution of the single reactions approximate in shape the normal bell-shaped curve. Assuming for a moment that premature reactions are included in the distribution, then the lower end of each curve would tend to weight the mean in proportion to the number of premature reactions included. It has been stated previously that premature reactions were made in the ratio of 1 : 3 : 6, in the normal, the incentive and the punishment series respectively. Then, of course, that is the cause of the improvement noted as being due to the incentive and the punishment factors. But, continuing the assumption, what happens if the lower part of each curve is cut off? By such an operation, the premature reactions would be excluded and if they had been the cause of the lowering of

the time in each series, then the remaining portion of each curve ought to be about at the same level. Obviously this does not happen. Even when the lower portion of the curve is cut off the remaining cases for each type of reaction group themselves on distinct levels with practically no overlapping. That is, the normal has a mode of 140, the incentive a mode of 130, and the punishment series a mode of 110. Furthermore, the punishment curve shows by its abrupt ending at the lower extreme that premature reactions were not included. Had they been present the curve would have trailed off more gradually and not ceased sharply. It shows also that the reaction-time of the punished reaction was very close to the physiological limit.

The evidence is strongly against assigning the improvement as caused by the premature reactions and the validity of the assumption that the incentive and the punishment are the factors reducing the reaction-time is unshaken.

CHAPTER IV.

CONCLUSION.

James[1] states, "The reaction whose time is measured is, in short, a reflex action pure and simple, and not a psychic act. . . . The preparation of the attention and volition; the expectation of the signal and the readiness of the hand to move, the instant it shall come; the nervous tension in which the subject waits, are all conditions of the formation in him for the time being of a new path or arc of reflex discharge. The tract from the sense-organ which receives the stimulus, into the motor center which discharges the reaction, is already tingling with premonitory innervation, is raised to such a pitch of heightened irritability by the expectant attention, that the signal is instantaneously sufficient to cause the overflow." In another place James continues, "Expectant attention is but the subjective name for what objectively is a partial stimulation of a certain pathway, the pathway from the 'center' for the signal to that of the discharge. . . . The signal is but the spark from without which touches off a train already laid. The performance, under these conditions, exactly resembles any reflex action. The only difference is that whilst, in the ordinarily so-called reflex acts, the reflex arc is a permanent result of organic growth, it is here a transient result of previous cerebral conditions."

This state was evident in this experiment in varying degrees under the several types of reaction conditions. Consequently premature reactions were sometimes made. The number that were made in each kind of reaction series may be regarded as an index of the degree of expectant attention in that series. From the total number of reactions obtained the ratio of premature reactions was found to be, 1 : 3 : 6, respectively for the normal, the incentive, and the punishment type.

On this point of attention Wundt[2] writes: "When we wait with strained attention for a stimulus, it will often happen that instead of registering the stimulus, we react upon some entirely different impression—and this not through confounding the one with the other. On the contrary, we are perfectly well aware at the moment of making the movement that we respond to the

[1] James, "Principles of Psychology", I., 90ff., 1890.

[2] Wundt, "Physiol. Psych.", 2nd, ed. II., 226.

wrong stimulus. . . . We cannot well explain these results otherwise than by assuming that the strain of the attention towards the impression we expect coexists with a preparatory innervation of the motor center for the reaction, which innervation the slightest shock then suffices to turn into an actual discharge. This shock may be given by any chance impression, even by one to which we never intended to respond. When the preparatory innervation has once reached this pitch of intensity, the time that intervenes between the stimulus and the contraction of the muscles which react, may become vanishingly small." To again quote from James[8], "Usually, when the impression is fully anticipated, attention prepares the motor centres so completely for both stimulus and reaction that the only time lost is that of the physiological conduction downwards. But even this interval may disappear, i. e., the stimulus and reaction may become objectively contemporaneous; or more remarkable still, the reaction may be discharged before the stimulus has actually occurred." In a footnote he gives an explanatory statement of the unusualness of such an occurrence.

With this background we may attempt an explanation of the effects observed in this experiment. One of the possible ones, and to the writer, the most probable is that a state of much keener attention, produced by an incentive and by the expectant punishment, motivates reactions very much prompter than those normally obtained under quieter and more restful conditions. The fact that there is a motive present stimulating the subject to react works the same in reaction-time as it does in all other types of responses to stimuli. How far this improvement may be carried on and to what limits the time of a reaction may be reduced is difficult of determination. The results show that the average time of the punishment series is about as fast as the individual is able to react and it seems the fear of punishment acts as a very powerful facilitator.

That a state of keener attention or concentration of the attention was involved is attested to not only by the premature reactions, but also by the feeling of the subjects. Especially during the punishment set did the subjects evidence the usual signs of close attention as well as copious perspiration and nervous tension. This fact is brought out in an introspective report of one of the subjects. "Introspections in connection with the punishment reaction. A continued feeling of tenseness through-

[8] Op. cit. I., 429.

out entire period of reaction while batteries were connected with key. At first, and at times during the fifty reactions, there was a certain shyness in the hand in approaching the key. This was especially noticeable after a shock had been received. When reacting it was impossible to inhibit impulse of withdrawing violently the whole hand, and the period of pressure on key was always accompanied by much instability or restlessness which often resulted in false or premature reactions. As the experiment progressed a feeling of intense warmth was present. All the above was in contrast to a more or less sleepy calmness in the preceding experiments without electric punishment."

GENERAL SUMMARY

1. Expectant attention reduced the reaction-time in the incentive and punishment series.
2. The factors of positive and negative incentive caused the state of keener attention to be maintained.
3. The limits of reduction depend upon the subject's training and upon the type of the individual, as the only limitation is the physiological time.
4. The actual saving, though not permanent, by incentive is six percent and by punishment is fifteen percent of the normal time. In terms of sigma the amounts saved are ~~eight~~ 9 and ~~twenty~~ 21 sigma respectively.

PSYCHOLOGICAL TESTS APPLIED TO FACTORY WORKERS

BY
EMILY BURR, Ph. D.

ARCHIVES OF PSYCHOLOGY
Edited by R. S. WOODWORTH
No. 55

Columbia University Contributions to Philosophy and Psychology

NEW YORK
May 1922

TABLE OF CONTENTS

CHAPTER I.

A Survey of the Psychological Testing of Factory Workers. 5

CHAPTER II.

The Factory and its specific Problems. 14

CHAPTER III.

The Tests. 19

CHAPTER IV.

A Discussion of the Reactions to the Tests. 22

CHAPTER V.

The Correlations. 61

CHAPTER VI.

Multiple Correlations. 73

CHAPTER VII.

Interpretation of the Data. 81

CHAPTER VIII.

Confirmation of the Tests:—Three Years Later. 84

Appendix. 87

Bibliography. 92

PSYCHOLOGICAL TESTS APPLIED TO FACTORY WORKERS

Chapter I.

A SURVEY OF THE PSYCHOLOGICAL TESTING OF INDUSTRIAL WORKERS

It is rather disconcerting that in the conduct of modern business the handling of the human element has received far less consideration than the development of machinery and the perfection of the materials used in industries. The history of industry shows a gradual increase in the specification of the material factors employed, but not until the last decade has the recognition of differences in the individual been considered.

In the industrial magazines of former years a great deal more space is devoted to the study of the materials and machinery utilized than to the analysis of the employee and the appreciation of his needs and value. When selecting a machine, managers of industrial organizations are extremely careful to obtain one that is absolutely adapted to the specific work for which it is required, but although there is more variation in men than in machines, in the great majority of cases men are still selected by the most haphazard methods. It should be obvious that even though all factories producing a similar type of goods were to adopt the same kind of mechanical devices their output would vary tremendously because of the differences in the quality of the human labor employed.

Until recently, neither the employer nor the employee gave much attention to the method of work. The employer interested himself first of all in the Output and the wages he was obliged to pay while the employee concerned himself but little with the matter of Output and a great deal with the matter of wages. The attitude of both toward the amount of energy that must be expended to bring forth results was similarly one of more or less indifference.

There is consolation, however, in the fact that among progressive employers there are those who are becoming aware that the management and selection of employees is one of the most essential and at the same time one of the most complicated details with which they have to grapple. Unfortunately, few have known how to go about the curtailing of the waste which invariably results from chance selection and the large labor turnover which inevitably follows.

In times of business stagnation, there has existed a surplus of applicants for industrial positions and the employer has been rather undisturbed regarding changes in his personnel. During seasons when the wheels of industry are moving at maximum speed, the undesirables have had to be employed as well as the trained and efficient. In normal times, however, when the selection of employees fitted to their task becomes a factor that influences the cost of manufacture, the employer lays greater stress upon scientific management and is more ready to lend an ear to the theory of proper selection of labor and to the methods and tests designed to determine its fitness or unfitness.

Psychologists who have observed the incapacity of workers who are suffering from some defect of the brain have tried vicariously to call the attention of employers to the futility of expecting certain results from individuals fundamentally unsuited for the tasks assigned to them, but although there is here and there an inclination to apply scientific methods in the selection of employees, as yet the field is being cultivated more in the pursuit of an ideal than in the interest of industry.

In order to judge of the fitness of an individual for a specific task, it is necessary to analyze not only the aptitude of the applicant for the work, but also the character of the work as well. Certain industries demand special dexterity or a degree of mental alertness quite at variance to that required by other concerns.

The importance of determining the aptitude of a man or woman for a particular job is thus being gradually recognized, but the industrial world has a long journey to travel ere it reaches the same efficiency in this direction as that which characterizes its use of machinery or of raw materials.

Prof. H. L. Hollingworth recommends that we make psychographs of the different jobs and determine the physical. psychological, intellectual, moral, social and temperamental character necessary for each. (1). This is a task to be undertaken in every trade and in every branch of the different industries so that the mental factors involved in each type of work may be analyzed and the varying degrees of ability which each demands be decided upon. An immense field for investigation of this sort lies open to the Experimental Psychologist and offers an opportunity for an almost limitless amount of research.

Among the first to recognize the importance of scientific research along lines that would aid in the industrial situation were Dr. J. Mc. K. Cattell (2) : Prof. H. Munsterberg (3) ; Prof. H. L. Hollingworth (4) ; Prof. E. L. Thorndike

(5); Prof. G. M. Whipple (6); to whose work reference is made in this paper.

In 1910, a Vocational Bureau was organized in Boston for the purpose of studying the human factor in industrial occupations. This seems to have been the first definite effort to analyze trade conditions from the standpoint of both employer and employee. It was found that only incidental attention was given to the subject of personnel organizations. In general, employees were treated only as economic units, and because of the lack of co-operation and understanding between employer and employee the labor turnover was enormous. The Bureau found that only a few firms scattered over the country had separate departments for hiring, training and promoting the welfare of their employees and that no scientific attempt to place applicants in positions suited to their individual capacity was being made by any of the firms then investigated.

A year after the inception of the Bureau, fifty men who were in charge of the Help Departments in fifty large establishments were asked to meet and discuss the problems of th employee. As a result of this conference, a concerted effort was made to apply the knowledge of the Twentieth Century towards the improvement of industrial practices. (7).

One of the chief factors which has brought about a more enlightened attitude toward the scientific examination and analytical study of the man or woman for the job, is the success attending the application of psychological tests to army recruits. During the war, the aid of psychologists was enlisted to examine and classify the intelligence of recruits so as to assign them to the duties for which they were best suited. The clever and ingenious tests then arranged proved a swift method of measuring the mentality of a man. Their application is some one million seven hundred thousand cases constituted a calculation of mental capacity on a larger scale than had ever been undertaken. (8).

A Committee on Classification of Personnel was appointed and it developed a number of Trade Tests which were successfully utilized in placing soldiers and sailors who came from every State in the Union. The tests were of three kinds: Standardized Questions, either oral or written, Picture Tests and Performance Examinations (9) which were arranged to determine a man's knowledge of the various trades.

The material for these tests was secured by a thorough investigation of each trade in order to discover suitable questions and comprehensive and fairly simple problems. When a series of questions was finally selected they were standardized by trying them upon groups of experts, journeymen, apprentices and novices. Questions which could not be answered, or work which could not be performed by novices

served to select the apprentices, while problems which could only be solved by the expert were intended to eliminate the unskilled. Those in charge of the Army Tests claim that their great value lies in their uniformity of method, in the brevity of time required to give them, and in the fact that they are so well standardized, that they can be given by persons without special training. There were two kinds of group examinations, one for men able to read and write English fairly well, called Alpha, and the other for persons not able to read and write English called Beta.

The interest awakened by these investigations of human intelligence has led to the discussion of the applicability of psychological tests in the industrial and business world. The question of the extent to which the present measuring devices of the army can be applied directly to industry depends, however, as Dr. Chapman so wisely observes, upon whether the conditions of an industry coincide with those of the army. (10). What the results of the Army Tests have proved is that it is no longer necessary to take a chance on a man's intelligence. Furthermore, they have demonstrated to the managers of business concerns a method whereby they may eliminate human waste and increase the efficiency of their employees. Therein lies the secret of their appeal to the business manager.

The testing that was done prior to that in the army did not arouse very much interest. Thus in 1915, when the American Tobacco Company engaged Prof. Scott to conduct some tests for that concern, there were only a few casual notices of the experiment. (11). The firm wished to have a scientific selection of salesmen made from a group of thirty-six applicants. It was arranged that six managers should interview the applicants and rate each by his own method. The results show the unreliability of haphazard methods of appraisal. The managers did not agree as to whether twenty-eight of the thirty-six applicants belonged in the upper or lower half of the group. One of the men was rated as Number One and as Number Thirty-two, another as Number Thirty and as Number Three, which shows the lack of uniformity in personal judgments.

John M. Bruce of the American Tobacco Company has applied certain standard tests that have proved useful in that industry. He claims that they determine "fitness, imagination, discernment, tenacity, address and personality." Each test is given a definite stated value, and the employee is rated according to his deviation from the average standing of the group. (12).

Thirty of the Efficiency Experts employed by the Cheney Silk Manufacturers were examined by Prof. Scott and the results of the tests correlated with the rating made by their

supervisors gave an index of plus .87. Cheney Bros. now use a series of psychological tests when interviewing applicants for high grade clerical and executive positions, or when contemplating promotions. (13).

The firm of Joseph & Weiss rely in a large measure upon the results obtained from psychological tests in their selection of both executives and operatives. They especially emphasize the value of the Directions Tests and state that these tests check up accurately in nearly every case with the status of the men in the ranking made by their bosses. (14).

The Dallas Consolidated Railway Company (15) used a series of tests which was composed of questions and a device designed specifically to test the powers of observation of the workmen. It consists of a contrivance representing a trolley track and various obstructions to traffic. The test is based on Prof. Hugo Munsterberg's method for selecting Sea Captains and is supposed, in this case, to determine a man's ability to think and act with speed and good judgment in an emergency. They use also a set of cards which are to be sorted according to the letter which predominates.

These methods of investigation have proved so valuable in the selection of their employees that the length of service of all train-men has been increased seventy-seven percent in less than five years.

The Curtis Publishing Company also apply tests arranged by Prof. Munsterberg as supplementary evidence in determining the efficiency and accuracy of their stenographers and clerks. (16).

The Bell Telephone Company have used Prof. Munsterberg's tests to determine the degree of dexterity, memory, attention, accuracy and intelligence possessed by would-be employees, and report satisfactory results. (17).

In his excellent book entitled "Hiring the Worker" Roy Wilmarth Kelly discusses various applications of psychological tests. Mr. Kelly found that eighteen of the thirty firms to whom he sent questionnaires had separate departments for hiring employees and that all save two had been established about 1912. Each of these Employment Departments had analyzed the tasks demanded of the worker, had recorded findings, and were endeavoring to solve the problem of the labor turnover. He states that only three of the firms were using psychological tests. One firm was much interested in the results he had obtained from three tests. Another firm had used tests to avoid hiring those who were mentally unfit, and another firm had found tests useful in preventing the promotion of the mentally incapable. In all these cases the tests were employed not as vocational guides but to discover the limitations of the individuals. One firm, to be sure, was formulating some psychological aids to assist

in the selection of employees. They were using a General Intelligence Test, a Card-Sorting Test, a Word-Building Test, a Trabue Completion Test and a Cancellation Test. No data regarding results or methods was given. (18).

Some good work in the application of scientific methods to sales-management has been done by St. Elmo Lewis who lays particular stress upon the psychological elements involved. He criticizes the fact that too few managers analyze with precision both the job and the ability of the man who is to execute it, and urges the necessity for the two operations. (19).

Wm. F. Kemble in his recent book, "Choosing Employes by Test," sets forth very clearly that painstaking analysis is needed to differentiate between men who are engaged in even as simple matter as lifting weights, and shows how time may be saved by determining which men are adapted to handling specific degrees of weight. (20).

One of the most comprehensive studies of the use of psychological tests under working conditions in a representative industry is to be found in "Employment Psychology" by Henry C. Link. The first half of this book deals with the application of psychology to the selection of employes; the second half deals particularly with its application to the training of new employes, the keeping of records of their activities, and a discussion regarding methods of transfer and promotion. The entire trend of the work described is a scientific attempt to attain the applicant's view-point. He urges the use of tests which will discover the applicant's potential skill, his innate ability, and his general intelligence and the amount of his experience. He gives accurate directions for the presentation of a number of tests and the method of marking and recording results. (21)

In an article in "Industrial Management" for April 1920, Dr. Spaeth tells a story of a manager of a large plant who, as the result of the presentation of tests for general intelligence, discovered among his employes many with "unsuspected brains"! These fortunate "finds" he immediately advanced to the benefit of both factory and the individuals. Dr. Spaeth quotes the results of a few tests made on a large number of girls engaged in the inspection of shells for rifle cartridges. The "inspectors" who were recommended because of their excellent standing in the tests worked much more intelligently and for longer periods than did those girls who received marks below the average in tests. (22).

Myer Bloomfield has analyzed recently the routine duties of the employes in one of the largest and most progressive firms in Newark, New Jersey. Changes have been made in

the firm's organization as a result of his study of both the employes and the executives. He analyzed all the different tasks of the concern and arranged tests to fit each one. Dr. Bloomfield is now connected with the Vocational Bureau of Boston and is carrying on work there of a similar character in a larger field. (23).

During the year 1917, F. F. Taylor of the Metropolitan Life Insurance Company formulated a questionnaire which he presented to the employes. The results give an insight into the individual characteristics of the men, and proved advantageous in reorganizing the personnel. (24).

Mons. Jaques has excellent success in choosing typists and stenographers by means of pschological tests and reports high correlations with the subsequent standing of the girls in their actual output. (25).

In the Manhattan Trade School a series of tests is administered to see whether a girl has sufficient mental capacity to make her intensive manual training advisable. The Woodworth & Wells Directions Tests and the Easy Opposites are among the tests usually given. (26). Dr. K. Murdock has completed a Sewing Scale by which the work done by these girls may be graded.

Some of the findings of the Committee on Employment appointed by the National Association of Corporation Schools are of interest because of their reference to psychological testing. In their report for 1919 it is stated that "the number of instances in which tests have been successfully applied warrants the belief that their use can be extended to other industries." They announce also that the "purpose of psychological tests is conservative and their technique sane," and add that the success of any application of them depends wholly "upon the seriousness and thoroughness with which the testing is done." (27).

Any resume of the application of psychological tests to workers in industry would be incomplete if it did not include some account of the tests conducted by Dr. Helen T. Woolley in the city of Cincinnati. To quote from her own statement, she says, "The Bureau of Vocational Guidance has for its ultimate object the comprehensive study of the problem of child labor with a view to furnishing scientific demonstration of the effect of labor upon children who enter industry early." The children tested were a group of eight hundred boys and girls, over fourteen years of age who came to the Working Certificates Office during the year 1911-1912 to secure working papers. They were tested in serial order whenever they chanced to apply at the office. The scheme made provision that each of the children should be retested mentally and physically each of five succeeding years. (28).

In connection with the study made by Dr. Woolley, the Chamber of Commerce of Cincinnati, under the direction of Supt. H. C. Storm, has undertaken a survey of the general industrial situation of that city. He has organized the Public School System so that it may co-operate with the shops especially in reference to the printing industry which is an important one there. The work is still in an experimental stage, and the results have not been published.

Bernard Muscio in his "Lectures on Industrial Psychology", discusses the question of the desirability of the application of psychology to industry. He demonstrates that by adapting psychological methods to industry it has been found possible to obtain a given output from a smaller expenditure of human energy than has hitherto been found necessary, and he urges Psychology to turn to its laboratories again in order to better what has been done and devise practical ways for furthering its application to industrial work. (29).

It seems probable that much of the social unrest in the world is due to the misfits in the industrial world. The worker's lack of interest in the task assigned to him, and often almost total inaptitude for the work makes him a drifter. The concensus of opinion on industrial conditions is that it costs from fifty to two hundred dollars to train a person for a position. Each individual manager is well aware of the costly and discouraging features involved in the constant changes of his personnel. But many of these managers must still be convinced that, at least to a large degree, this bugbear of all industrial management, the labor turn-over, can be cut down by the exercise of greater care in engaging employees who are suited to their jobs.

The history of Vocational Psychology has been one of trial and error, and it needs must be conducted empirically for a long time to come. In its very essence it is made up of variable elements for some jobs change character almost from day to day, and the instability of the human element is readily conceded. In any review, then, that one makes of the subject of psychological testing one should note that the Vocational Tests are still for the most part in the experimental stage. There are no cut and dried methods that apply to all industries alike. Each one must solve its own problems. (30).

Tests have been standardized successfully in only about seven special lines of work other than "trades." The tests for the "trying out" of typists, stenographers, correspondents, labellers, clerical workers, inspectors, seamstresses and salesmen have all proved satisfactory and the norms established in each are reliable. (31).

The Trade Tests that were being standardized for use in the Army under the direction of The Committee on the Classification of Personnel have not yet been published. They were devised for some seventy-five or a hundred different types of trades and will doubtless be of greatest value when given to the public in the interest of vocational guidance. Dr. Chapman has described certain tests applicable in about thirty trades, and in his book entitled, "Trade Tests" gives accurate directions for their use. (10).

The study described in this paper was undertaken to demonstrate the value to be derived from the application of psychological tests in a hitherto unexplored field of industry. In addition to performing this function, the investigation is destined to give us information regarding the general intelligence of a certain group of women and girls who are earning a livelihood in industrial work. The mentality of delinquent women has been subjected to painstaking examination, (32), but the mental development of women factory hands, none of whom have failed to adjust themselves to social requirements, has received inadequate attention.

Chapter II.

THE FACTORY AND ITS SPECIFIC PROBLEMS.

Every time it is definitely proved that the reactions to any test, or to any group of tests, provide a means for the measurement of the capacity of candidates for a certain type of industry, a distinct service is rendered, not only to the employer, but to all potential employees.

The Experimenter secured the co-operation of E. Eisemann & Company, owners of a Feather and Fancy Ornament Factory situated at 63 West 38th Street, New York City, and received permission to submit their employees to a series of psychological tests.

Several days were spent in observing the work of the factory and studying the operations in the different departments. The Experimenter, although not actually serving at the bench, tried out each type of work herself so that the analysis might be as accurate as possible. Various members of the firm, as well as the heads of departments and the factory hands themselves, were interviewed and the Experimenter was able to profit by many of the suggestions they made.

The Manager said that he had relied largely, when engaging employees, upon the impression made upon him by the general appearance of the applicant and her replies to his questions, but he added that he would save time if he could have a more scientific method on which to base his decisions. He recognized the fact that his method did not permit him to discover natural aptitudes and he therefore formed his concepts of character and ability from observation alone. He would be glad, he said, to utilize a series of tests by which he might sort out applicants for employment, or use when promotions were to be made.

Before analyzing the characteristics which employees in this factory should possess, a careful study of each type of work had to be made.

The factory is divided into six distinct departments as follows:

FANCY DEPARTMENT
SELECTING DEPARTMENT
BENDING & FINISHING DEPARTMENT

BOA DEPARTMENT
DYEING DEPARTMENT
ERRAND DEPARTMENT

The work in the "Fancy Department" consists of sewing, glueing and assembling in endless combinations all sorts of feather materials in duplication of a presented model. This necessitates the exercise of judgment, discrimination, the ability to follow instructions and dexterity both with the needle and with the fingers. If a girl be imbued with imagination, initiative and creative ability, she is encouraged to originate models of her own design. In general, a higher order of mentality than that required in the other departments is demanded of the workers in this field, and consequently these girls receive the highest salaries.

The greatest number of girls is employed in the "Selecting Department" where the task consists of sorting the feathers according to their length and width. This requires the exercise of the powers of concentration and observation and demands a rapid visual estimate which must be accurate. A fair degree of discrimination is needed. This work is of a monotonous nature and involves the continual exercise of the same muscles of the body and of the same elements of the mind. Since the work is monotonous and not particularly interesting the labor turnover is greatest in this department. The work is divided into two sections, one which the raw feathers are sorted and the other in which the dyed, finished product is assembled.

The "Benders and Finishers" are engaged, as the designation implies, in the work of forcing the feathers into certain predetermined shapes. Manipulative skill, a great deal of patience and sufficient intelligence to follow instructions are needed. Attention is the most important factor upon which demand is made.

In the "Boa Department" the work requires the same dexterity and persistency but under more trying circumstances since the bits of feather to be assembled are very small and accuracy in manipulation must not be sacrificed to speed. Girls of a nervous temperament are sorely tried by work of this character and rarely accomplish the same amount of work with the same amount of exactness as do fellow-laborers blessed with a more phlegmatic disposition.

The "Dyeing Department" employs more men than women and its secrets were not divulged beyond the fact that some knowledge of chemistry and a fair degree of intelligence are necessary requisites. Only two girls (167 & 117) from this department were tested.

In the "Errand Girl Division" the work consists mostly of fetching and carrying from one department to another.

The youngest and least experienced girls do this work, usually serving an apprenticeship as errand girls before being transferred to one of the other departments. It was, therefore, deemed of interest and importance to give them tests to determine for which type of factory work they were best suited. As Errand Girls per se, a fair memory and a spirit of willingness and co-operation as well as good comprehension are obviously the most essential characteristics demanded.

Racially, the workers might be divided, either by birth or descent, into four fairly distinct groups consisting of Austrians, Irish, Italians and Russians. The factory manager stated that he endeavored to keep these groups more or less distributed so that there would be slighter danger of labor union interference and the disturbance which would of necessity be caused when Jewish and other holidays occurred.

In reply to questions regarding their attendance at school, thirty-two of the subjects said that they had been graduated from the Public or Parochial Schools of New York City.

One girl had left school when in the Third Grade; three when in the Fourth Grade; two when in the Sixth Grade; seventeen when in the Seventh Grade; and eight when in Grade Eight. The remainder "could not remember" what grade they were in when they last attended school. Although the Experimenter was unable to verify these statements, there is no reason to doubt that they are in the main truthful.

TABLE ILLUSTRATING ABOVE SCHOOL HISTORY

Graduated of P. S. or Parochial Schools	32
Left school when in grade indicated,	
Third	1.
Fourth	3.
Fifth	2.
Sixth	7.
Seventh	17.
Eighth	8.
Did not recall	5.
Total No. of subjects	75.

When asked at what age they had last attended school, the replies indicate that the average age of those who left school before they were graduated was 14.5 years and the average age of those who did graduate was 13.73 years at the time of graduation. This suggests that the first group is composed of persons of a slower and less keen mentality than is to be found in the second group.

Many of the subjects regretted their lack of education yet said that they had been able to do their factory work fairly well and they seemed to like it in spite of its rather limited possibilities, that is they probably preferred work to school. They had all left school without having acquired any definite industrial or business training and it was by mere chance that they had drifted into this type of work. An introduction by relatives or friends already in the factory was the decisive factor. There had been no consideration of the suitability of the person for the job.

The subjects ranged in age from fourteen to fifty, but the majority were between sixteen and twenty-one years of age, as is indicated in this table:

Ages	14	15	16	17	18	19	20	21	22	23	24	25	26	36	40	50
No.	1	3	5	15	12	7	12	4	4	6	3	3	1	1	1	1

The spirit of the employed force was wholesome and pleasant. The Employer provided sick benefits, a service of free umbrellas and rubbers on rainy days, a small circulating library, and at more or less frequent intervals, social events which were arranged either by the firm directly or by the employees with the aid and support of the firm.

There were, therefore, no ulterior and difficult factors to be overcome before the testing was begun. No artifical or abnormal conditions that required adjustment seemed to exist.

While analyzing the jobs, the question of what tests to use was borne constantly in mind. There should be, it it seemed, an estimate of the level of intelligence from the standpoint of the individual and of the various groups engaged in the performance of the different types of work.

The Management emphasized especially the importance of ascertaining the capacity of a subject to comprehend and carry out instructions, so Directions Tests must be given to measure this particular ability.

In addition to these estimations, the perception, memory, judgment and reliability of the employee should be determined. She should be tested also as to her aptitude for making eye-hand co-ordinations. Her steadiness of control and concentration when a monotonous task is demanded of her should by all means be measured.

The degree of dexterity she possesses, her sense of color and proportion should in some way be demonstrated by the test-results.

The tests eventually chosen are described in the next chapter. Just what distinctive mental traits are denoted by an ability to perform this or that specific test cannot be definitely determined. There is too much over-lapping of mental qualifications to make this possible, but we can, in a measure, suggest some of the most important abilities demanded in the execution of the different tests.

Chapter III.

THE TESTS.

In order to establish test conditions as nearly standard as possible the co-operation of practically every one in the factory was sought and readily obtained. In addition, a spirit of keen interest in the tests and of friendliness toward the Experimenter was created which aided materially in the testing. The reasons for the investigation were explained to each one in turn. There was difficulty in gaining the confidence of but one of the subjects, an older woman, who felt that she was about to be submitted to some indignity that reflected upon her age and long connection with this business. She became interested, however, before she had progressed very far with the tests, finally confessed that she used to love school and wished she could have "gone higher." In the testing, therefore, little need be discounted on the score of nervousness or a lack of co-operation.

The seventy-five subjects who were examined by these tests and whose records are discussed later, spoke English readily and possessed a vocabulary, which if limited in scope, was yet used with perfect understanding and significance.

The Manager permitted the employees to leave their work for the tests, and since no recreation time was encroached upon, they were grateful for this break in the dull monotony of their daily tasks and reacted in a cheerful manner. In several cases, subjects expressed a positive delight in the tests which one of them described as "grand fun."

The tests were conducted in the Rest Room of the Factory and extended over a period of many weeks. The workers were examined singly, a proceeding that is probably unprecendented in the history of the testing of factory hands.

Before the actual testing was begun, the workers were made familiar with the method of examination by a series of preliminary tests the results of which were neither recorded nor used. This was done largely to eliminate the initial nervousness which fear of the "unknown" might have instilled in the hearts of some of the girls.

Each subject was given an identification number which appeared on all her test sheets. To distinguish between her identific number and the order given her in the firm's ranking, 100 was added to each identifying numeral. Thus

all numbers over 100 refer to the examiner's classification, while all numbers below 100 refer to the firm's classification.

A stop watch was used and the time required for the performance of each test was recorded.

The Experimenter found as did Dr. Link, (21), that unless the stop watch is concealed from the subject's view, she is inclined to be nervous and distracted by the idea that she is being timed. Only in certain tests were time-limits enforced. Limitations as to time would have defeated, in a measure, one of the secondary aims of this investigation, namely, to determine how the time records of these girls who are all engaged in factory work, compare with those of other types of subjects.

The directions were given orally because it was found that the girls follow spoken instructions much more successfully than written ones. The test sheets were kept face down until the subject was directed to begin.

Instead of devising many new tests, it seemed wise to employ tests in which norms could be used for comparative purposes. No detailed description of these tests is necessary inasmuch as they are widely known. The new tests which were arranged solely for use in this particular type of industry are discussed under the Sixth Group of Tests.

The mode of presentation and of scoring these Special Tests is in the Appendix.

The following is a list of the tests, presented in the order in which they were used:

I. General Intelligence — Binet-Simon
(Goddard Revision.)

II. Association Tests — Woodworth & Wells
A. Easy Opposites.
B. Mixed Relations No. 1.
C. Mixed Relations No. 2.
D. Substitution.

III. Cancellation Tests.
A. Digits Cancellation.
B. Letter Cancellation.

VI. Linguistic Tests.
A. Trabue Completion.
B. English-African Proverb Test-Ruger.

V. Directions Tests.
A. Easy Directions No. 1.
B. Easy Directions No. 2.
C. Hard Directions.

VI. Special Tests for this Industry.
- A. Judgment Test.
- B. Card Sorting Test.
- C. Sorting Test.
- D. Discrimination of Color Test.
- E. Motor Control Test.
- F. Feather Sorting.
- G. Weight and Form Test.

SPECIAL TESTS FOR THIS INDUSTRY.

These tests were arranged to call forth the different types of activity demanded in the several departments of the Factory.

Since the Selectors must judge feathers from the viewpoint of length, width and color as well as quality and suitability, Tests A., B., C., D. and F. should be especially desirable to aid in the discrimination of workers in this department.

The Boa and Stringer Department requires some ability to form judgments and a great deal of motor control, hence Tests A. and E. should be found useful indices of the best workers in this department.

For the Fancy Department Tests A. and D. seem to be indicated particularly.

The tests that require discrimination and quickness of perception may very profitably serve to demonstrate the degrees of proficiency attained by the various employees.

The Sorting Tests and the Discrimination of Weight and Form Tests should reveal the manual dexterity possessed by the girls.

The determination of the general intelligence of a girl, and her ability to comprehend and carry out instructions given her have been left to criteria established by the tests in which norms have already been ratified. The correlations derived from a comparison of the standing of the members of this group in the tests and their rating in the work of the factory is described in detail in Chapter Five.

The application of several tests has been employed in each case rather than the arbitrary presentation of a limited number of tests. Ability in one test compensates for deficiency in another, hence to present tests requiring different mental functions is the fairest way to form an estimate of the individual's capacity.

Chapter IV.

A DISCUSSION OF THE REACTIONS.

I. The Binet-Simon Test.

This test as arranged by Dr. Goddard (33) was presented in the prescribed manner to each of the seventy-five subjects.

The mental ages of the individual in the group range from 8.6 to 12. with the crude mode at 11.2. Twenty-nine fall below the mental age of 11 and only two attain a rating of 12. One employee passed but one test, (X-1), above the 8th year, and two others received a mental age of 9.4 and 9.6, respectively. The girls who had the lowest grades of mental development were found, with two exceptions, to be working on the boas, or bending feathers, or sorting them.

The distribution of mental-age on the Binet Scale may be observed by an inspection of this table in which the mental age is indicated and the percentage of subjects obtaining each age is given beneath it. The Median Age is 11: S. D. is 1.1358.

Mental Age	12	11.8	11.6	11.4	11.2	11	10.
% of Subjs.	3%	12%	11%	9%	18%	8%	11%

Mental Age	10.6	10.4	10.2	10.	9.6	9.4	8 8
% of Subjs.	11%	8%	3%	3%	1%	1%	1%

It is interesting to note how the results of this group, measured by the Binet-Simon Scale, compare with the records made by other groups.

Dr. Katherine B. Davis (34) in her Study of Prostitution in New York City writes that among 647 women at the Bedford Reformatory examined by the Binet tests, there were but two who measured 12 years mentally, twenty-six measured 11, and forty-four obtained a mental age of 10. The rest were below 10 mentally.

In this Bedford group, one-third of one percent had a perfect score compared to 2.66 per cent in our factory group; and while only four percent of the former group ranged in mental age from 11 to 11.8., inclusive, fifty-eight per cent. in

our group reached a mental age ranging from 11 to 11.8 Furthermore, whereas, 6.8 per cent reached a mental age of 10 to 11 in the Bedford Group, thirty-four percent received a like grading in the factory group. We may conclude, therefore, that this selected group of factory girls is of a higher type of mentality than the delinquents. Their opportunities along educational and social lines have been similar and one is led to deduce from a comparison of the results of these tests that the radical differences in the two groups are due in large measure to variations in native capacity.

The Binet-Simon tests conducted by Dr. Olga Bridgman and Dr. L. Morrow at the State School for Girls at Geneva, Illinois, show that but six of the girls are capable of passing the 12-year Test. Fourteen of the sixty tested were retarded from one to three years, and the remaining forty were retarded more than three years. The girls were all over sixteen years of age. That is eleven were retarded from four to five years and twenty-nine from six to thirteen years. (35).

In giving the Yerkes adaptation of the Binet-Simon Test to a group of mill-operatives, Dr. Yerkes found that errors were committed in all but three of twenty tests. In a Y. M. C. A. group, to whom the same tests were given, errors were made in all but five of the tests. Both of these groups consisted presumably of persons who had as good or better schooling than had the factory girls, yet errors were made in nearly every test. One can make no exact comparison between the Goddard revision and the Yerkes adaptation of the Binet, but this notation seems of interest inasmuch as basically the two tests are alike. (36).

In the "Report on Sensory and Mental Tests" made at the Idaho Industrial Training School, the author states that no attempt was made to apply any tests above the 12-year Scale of the Goddard Rivision of the Binet-Simon Test. All subjects who passed the 12-year Tests were counted normal. Eighty-three received a rating of 12., while fifty-five tested below twelve and were consequently called sub-normal. (37).

In his results obtained from a study of "The Measurements of Delinquents," T. H. Haines (38) states that among the one thousand individuals to whom he gave the Binet-Simon Test, five hundred sixty-three were rated below the 12th Year. In this connection, he speaks of the fact that some examiners, who use this scale, class those who measure below 12 as Feeble-minded, and he proceeds to disprove the method of classification that arbitrarily places the standard of demarcation at 12 years.

The rigid application of such a rule as this would place all save two of the Factory Group tested in this experiment in the sub-normal category, an obviously absurd classification.

Both Prof. E. L. Thorndike (39) and Dr. Stern (40) have shown the inaccuracy of the Binet rating above the age of 9 Years, and it must be conceded that to place the line of demarcation between the normal and the Feeble-minded at the 12-Year level on the Binet Scale is arbitrary and scientifically questionable. The amount of retardation that makes it impossible for an individual to manage her affairs varies with the individual and is to a large degree dependent upon her environment. In factory work, a girl of high intelligence is not necessarily the most desirable. One may note in this investigation that girls who have a mental age of 10 years, or thereabouts, are making good in this factory.

Let us next consider the individual tests of the Binet Scale, ranked in order of difficulty, beginning with the one in which most errors occurred.

This proved to be the repetition of sentences (XII.-3), in which 87 per cent. of the subjects failed. The dissected sentences came next with 62 per cent. incorrect; while 56 per cent. of the girls yielded to suggestion; 47 per cent. could not repeat seven digits (XI-1) in two out of three trials; 47 per cent lacked sufficient visual memory and analytical skill to reproduce the two designs (X-2); 44 per cent. were unable to discriminate between small differences in weight (IX-5); 29 could not repeat six digits two out of three trials (X-3); 12 per cent. failed to give definitions of "Charity," "Justice" and "Goodness" (XII-2); and 12 per cent. were unable to construct a sentence to meet the requirements of the Tenth Year (XI-5); 10 per cent. were unable to answer the Comprehension questions (XI-4); 9 per cent. gave the definitions of simple objects in functional terms only (IX-2); 6 per cent. did not give sixty free associations in three minutes (XI-3); 5 per cent. of the girls did not see the absurdities (XI-1); 2 per cent. could not repeat the names of the months without more than one error (XI-4); and one person (151) could not make change.

It has been stated that the subject finds the Binet test difficult in proportion to the time that has elapsed since his formal education ceased, but our results reveal nothing conclusive on this point. There are as many failures, comparatively, among those fresh from school as among the older girls and women who have been out of school many years. School Training to judge from an inspection of the table given below, has not influenced to any extent the rating on the Binet Scale of the members of this group. In this table, the employees are arranged on the basis of their grade on leaving school. The failures in the individual Binet Tests are indicated and the range in mental age of each group of those who left school when in a like grade are noted. It may be observed that the type of failures do not differ much from one group to

another. Thus, some public school graduates have a mental age of 10.2 while other subjects who left school from the 4th., 5th., or 6th. grade have received a mental age of 10.2; 11.2; 11.8 etc.

This peculiarity is in keeping with results obtained in the testing of other groups, notably the Army Group, where it was found that some of the highest intelligence records were made by men who stated that they had never completed the Elementary Grades.

It seems, therefore, that we are justified in concluding that the Binet-Simon Test is a measure of General Intelligence and native capacity which does not depend to any great extent upon school drill.

TABLE 3. THE BINET SIMON TEST RESULTS ANALYZED ON THE BASIS OF SCHOOL ATTENDANCE.

Group I. The thirty-two subjects in this group are graduates of Public or Parochial Schools. Their mental ages range from 10.2 to 12. The different tests in which these subjects failed to receive credit are noted below under the special test indicated:

IX	X				XI					XII					(Mental Age).
5	2	3	4	5	2	3	4	5		1	2	3	4	5	(Test Number).
4	8	5	2	1	6	1	1	11		12	1	18	15	7	(No. of Subjs.).

Group II. The seven subjects in this group left school in the Eighth Grade. Their errors are indicated as above. Their mental ages range from 10.6 to 11.8.

XI	X		XI					XII	(Mental Age)
5	2	3	1	2	3	4	5		(No. of Test).
3	6	1	2	2	6	6	2		(No. of Errors).

Group III. The nineteen subjects in this group all left school when in the Seventh Grade. Their errors are indicated in the same manner as are those of Grade I. The mental ages in this group range from 10.2 to 11.8.

IX	X				XI					XII					(Mental Age).
2	3	4	5		3	5	1	2	3	4	5				(No. of Test).
8	5	3	1		2	12	12	11	14	17	6				(No. of Errors).

Group IV. The twelve subjects who left school when in grades ranging from the Third through the Sixth Grade. Their mental ages range from 8.6 to 11.6., inclusive.

IX	X				XI				XII					(Mental Age).
5	2	3	4	5	2	3	4	5	1	2	3	4	5	(No. of Test).
7	6	3	3	3	2	5	1	7	6	1	9	5	2	(No. of Errors).

Group V. Five subjects could give no definite information regarding their school history. The mental ages of these subjects range from 9.6 to 11.8.

IX	X				XI					XII					(Mental Age).
5	2	3	4	5	1	2	3	4	5	1	2	3	4	5	(No. of Test).
2	3	2	1			1	2	1	3	1	2	2	4	2	(No. of Errors).

Since the only Auditory Memory Tests used were those included in the Binet Scale, a brief discussion of their results is of interest. Fifty-seven per cent. of the subjects have at least a seven digits memory span, and of this fifty-seven percent, two subjects passed all the Binet Tests; two obtained a mental age of 10.4; two of 10.6; three of 10.8 and three of 11 ; the remainder were about evenly divided between the mental ages of 11.2 through 11.8. Those who had but a six digits memory span comprised thirty-two per cent. These ersons obtained the following ratings on the Binet Scale: hree per cent. were rated 9.4; 10.4; and 11.6, respectively; and six per cent. of this six digits memory span group had a mental ge of 10.2 and the same per cent. of the six-digits group ob-ained a mental age of 11.

The Five-Digits Memory Span Group consisted of eleven per cent. of the subjects, and they measured 8.6 (one only),

9.6 (one only), 10.4 (two only), 10.6 (two only), while the remainder were graded 10.8 and 11. mentally.

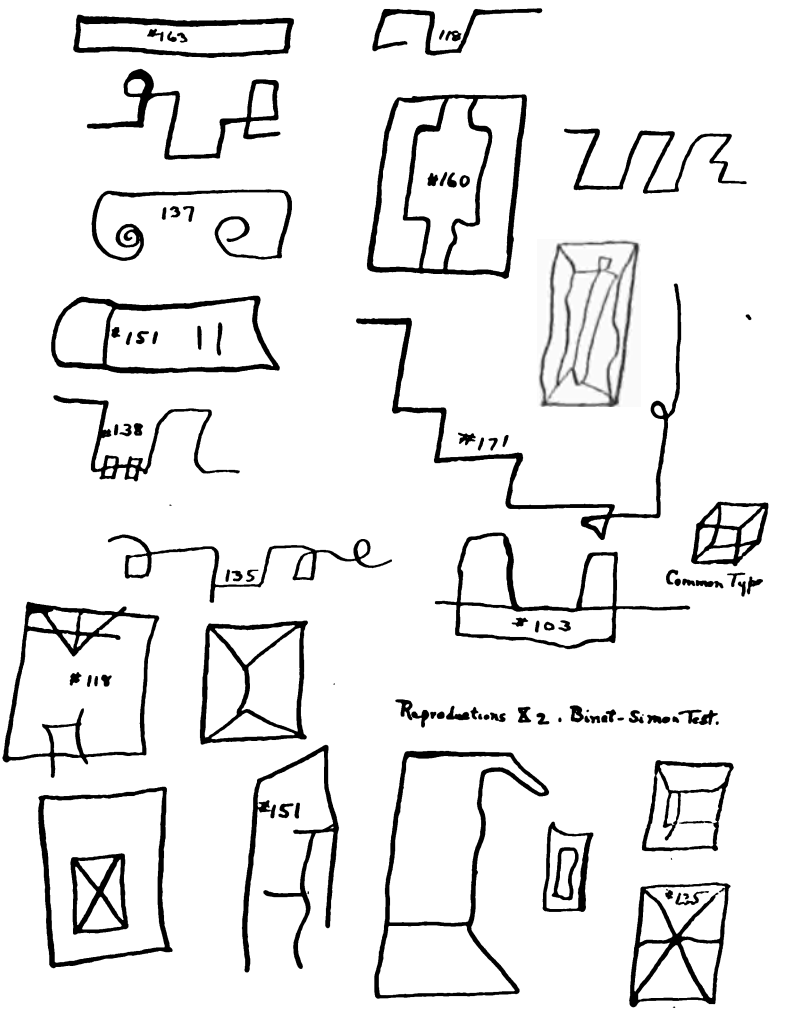
Thirty-eight per cent. of the Bedford Reformatory women have a Memory Span of not more than six digits; and thirty per cent. fail to recall even that number. (31).

In experiments with one hundred sixty-five college students, Gates (41) found that the average memory span for the group was 7.7 digits.

The result obtained from the auditory memory tests applied to this group of Factory Girls have been analyzed and given a Four Grade rating. Grade A denotes that the employe has at least a seven-digits memory span and has also been able to repeat after the Experimenter a sentence containing twenty-three syllables. Grade B signifies that the subject cannot repeat any one of the three sentences verbatim, but does possess the ability to repeat seven digits. Grade C shows that the subject's memory span is limited to six digits; and Grade D indicates that the subject has a five digits memory span only. This method of rating the auditory memory of a subject gives a basis for a fairly explicit type of group analysis that is useful in determining the concentration of employees in different departments of the factory.

Investigating still further the mental qualities requisite for workers in this factory, the reproduction of the two designs was found to have some "vocational value." It is an exercise that is quite on a par with the mental effort required from the workers in the Fancy Department in duplicating models submitted to them for reproduction. This is so true that those who failed utterly in this test were invariably found to be employed in other then the fancy Department.

The accompanying reproductions of these designs indicate how far astray some of the subjects went in visualizing the models presented.



It is obvious from an inspection of the following Table that employees who fail most decidedly in these tests are not disqualified for work in the Selecting, Boa and Bending Departments. It cannot be chance that no one employed in the Fancy Department has failed in the test.

Identification	Bank on Firm-Scale	Dept. of Factory
103	71	Boa-Stringer
107	35	Selecting
118	73	Boa
121	46	Selecting
128	69	Bender-Twister
135	45	Selecting
151	74	Selecting
160	67	Selecting
163	20	Selecting
164	18	Selecting
168	30	Selecting

WOODWORTH & WELLS ASSOCIATION TESTS.

I. Easy Opposites. (42).

This test was presented in the usual manner. The subject was asked if she understood what is meant by an "Opposite," and three illustrations were given her. She was then told to give the opposites of "Yes," "High" and "Wet." If the girl was unable to do this correctly, the directions were repeated and the words, "Big" and "Up" used as further illustrations.

In spite of these careful preliminary explanations, errors occurred which showed that the problem had not been properly digested. An analysis of the erroneous reactions reveals a failure to carry over an idea that in several previous cases had appeared to be perfectly understood, or else the inability to differentiate between words that sound alike yet have a different meaning and are spelled differently. There are still other instances in which an attempt at a definition has been made, and some reactions seem only explicable on the ground that associative processes were aroused by the stimulus word alone and the initial reaction was written without further thought.

"East" and "West" are reactions of this type, when given as opposites of "North." It is probable that had the subjects making these responses been catechized orally on this particular opposite each one would have known that South is the opposite of North. As an opposite of "Slow" one finds "Weak"; and as an opposite of "Weak" one finds "Thin," "Sick," "Brave" and "Well"; and evidently mistaking Weak for Week, some subjects gave "Day" or "Month." Attempting to define the terms, subjects gave "Tall" and "Big" as antonyms of "Large"; "Beyond" as an opposite of "Above,"

and "Delicate" as the opposite of "Good." Bua the mental proceses which prompted such reactions to "After" as "Now," "Seen," "Beyond," "Late," "Went," "For,"—are not easy to analyze. The opposite of "Love" resulted in many failures among which are "False Friend," "Mariage" and "Dont love you." "Come" brought forth such reactions as: "Went," "Going," "Stay," "Before," "Gone," "Don't be gone" and caused more errors than any of the other stimulus words.

No one neglected to give the correct opposites for "Rich" and "Dark," although they apparently are no less difficult seems probable that errors are due quite as much to a state of mental sluggishness as to a lack of linguistic knowledge. than some of the words that were frequently missed. It

The average reaction time for this group was 89.1 seconds; the average deviation in time was 25.92 seconds and the standard deviation in time was 34.75 seconds. The time range is from 29 seconds to 218."

The average score was 93.64 per cent. with an A.D. of 6.84 and S. D. of 30.45. The median score was 95 per cent. The range in score from 60 per cent. to 100 per cent. on the basis of 5 per cent. credit for each correct response. No half-credits allowed.

In the tests conducted by Dr. Bronner upon a group of delinquent girls, she reports a range of time in the Easy Opposites Test of 35" to 145", while in her Evening Settlement Group, the time range was from 37" to 130". In the Servant Girl Group, there was a time range of 52" to 215". The delinquent girls had no perfect scores, though some of the subjects missed only one-half of one opposite. (45).

The reports made by the Servant Girl Group in the Opposite Test show, according to Dr. Bronner, scores in which there were only 5., 6., 11., 12. and 13 correct opposites. In our group, the poorest performance was credited 60 per cent. or twelve opposites right, and 33.3 per cent. had perfect scores. Furthermore, Dr. Bronner permitted half-credits which has not been done in our scoring. While the entire lists of words were not exactly the same, they were of equal difficulty, approximately.

The probable reason for the prolonged reaction time of the Factory girls is due to the fact that they wrote their own responses. The girls in Dr. Bronner's group did not write their own reactions, and in the Easy Opposites Test conducted by Dr. Weidensall (32) at Bedford Reformatory, the reactions were written for the poorer half of the subjects. These discrepancies in method, unfortunately, make it impossible to offer any exact comparisons as to the time-records of these various groups.

2—The Substitution Test.

This test was presented in order to determine the rapidity with which new associations are formed by repetition. The practical problem is to combine the greatest possible speed with the smallest number of errors and consider both factors proportionately.

The time score is the chief guide because the mistakes are negligible,—six as a maximum, and a perfect score in the majority of cases. Yet our aim must be to find a formula which gives full value to both time and error and makes the material directly comparable in the form of one numerical value instead of two. In scoring, the method of penalties described by Prof. Munsterberg was employed. (46). By this method, the approximate range of both the quantitative and qualitative performance was found and then the units on the quantitative range were equated with the units on the qualitative range. The method equates time and speed, and by evaluating the results of the tests in this way a general conclusion as to the employees' efficiency may be made.

The average time score of the factory girls in this test was 176.5 seconds with an A. D. of 23.36., and a S. D. of 37.4". The time range was from 95" to 348.5".

Since the errors in this group ranged from 1 to 6, this makes the penalty for an error, 40.8" according to the method described above, and the employees were arranged in rank order of merit in the test on the basis of this penalization for errors.

When this test was presented by Woodworth and Wells to eleven adults the average time required for the test was 144.7 seconds, with a Standard Deviation of 12.5" and a range in time of 111 seconds to 176.5 seconds. (43).

The average time for the first half of the test in the Woodworth and Wells Group was 79.66 seconds in contrast to 92.4" required by the Factory girls to do a similar portion. The second half of the test was performed by the Woodworth and Wells Group in 65.1", while the factory girls required on an average 84.2" to do the latter half of the test.

Dr. Weidensall (31) reports a greater time range than that of the factory group. The Bedford women had a time range of 181.16 seconds to 618.5 seconds. The average reaction time for the first half of the test was 176.83 seconds and for the second half was 118.23 seconds. One is, therefore, justified in concluding that new associations are formed more slowly by the delinquents than by this particular group of factory workers, who, in turn, required a longer time and revealed a wider time range than did the adult group examined by Prof. Woodworth and Dr. Wells.

B. Mixed Relationship Test No. 1

Before presenting this test, the Experimenter assured herself that the subject understood fully what she was to do. Three examples were given her, outside of course, of those submitted in the examination; thus, Knee is to leg as Elbow is to? Shoes are to feet as Hats are to? Chew with one's teeth and Sew with one's? In this particular, a deviation was made from the regulation procedure. It will be recalled also that in this and in all other tests, directions were given orally. During the trial testing, the subjects appeared unaccustomed to written instructions, and asserted that they never made written lists of commissions to be remembered, and rarely, if ever, received other than spoken orders. This change in method seemed, therefore, to be wholly justifiable.

Inasmuch as this test requires more linguistic knowledge than the giving of Easy Opposites, the scores obtained are correspondingly lower.

The average score was 61.11 per cent.; the average deviation was 37.62. There were three perfect scores. The median score was 65 per cent. Scores ranged from 10 per cent. to 100 per cent. The Standard Deviation was 44.82.

The average time required for the test was 196.7"; the A. D. was 53.45; the S. D. was 70.73 seconds. The time ranged from 93 seconds to 400.4 seconds.

Woodworth and Wells give an average reaction time for a single response of 3.14 seconds, while the time required for a single test in the factory group was approximately 6.83 seconds. The subjects in the former group were college students habituated to intellectual exercises, and, therefore, scarcely comparable to the factory girls in a test of this kind. (44).

The errors that occurred in this test, analyzed, reveal an inability to adjust ideas in a new and unwonted manner. The failure to "carry over" a conception from one situation to another was, as in the previous test, the most frequent type of mistake. The repetition of one or another of the three words given instead of a reaction to the third word only often occurred. In a few instances, perhaps simply to avoid leaving a blank space, words that had appeared previously were written. Some of the reactions can be accounted for in no other fashion.

A few illustrations of the typical errors are given below; the words supplied are bracketted.

Eagle—Bird—Shark ..... (Bird, Snake).
Eat—Bread—Drink ..... (Starve, Glass, Eat).
Fruit—Orange—Vegetable ..... (Eat, Meat, Orange).

Sit—Chair—Sleep .....(Well, Walk, Lay, Lie, Rest, Night, Awake, Asleep).
Double—Two—Triple .....(Double, Single, Twice, Fish, Babies).
England—London—France .....(Germany, Italy, Spain, French, State, City, Country).
Chew—Teeth—Smell .....(Chew, Mouth, Gum, Odor, Flowers, Taste, Perfume, Gas, Good).
Pen—Write—Knife .....(Knife, Fork, Tool, Write, Sharp, Sight).
Water—Wet—Fire .....(Water, Out, Red, Sticks, Shrink).
He—Him—She .....(He, Him, Then, It, Woman).
Boat—Water—Train— ....Car, Engine, Elevator, Machine, Runs).
Crawl—Snake—Swim .....(Snake, Water, River, Bath, Swam, Walk, Float, Flew, Quick).
Horse—Colt—Cow .....(Colt, Kein, Dumb Animal, Bull, Pig, Dog, Milk, Chew).
Nose—Face—Toe .....(Finger, Face, Head).
Bad—Worse—Good .....(Bad, Best, Pure, Well, Excellent).
Hungry—Food—Thirsty .....(Hungry, Starved, Dry).
Hat—Head—Glove .....(Glove, Head, Shoe, Wear).
Ship—Captain—Army ....(Navy, Leader, Battle, Regiment, Men, Land).
Man—Woman—Boy .....(Woman, Male, Child, Men).

C Mixed Relations No. 2

This test appeared as No. 1. in the Monograph (14), but is here designated as No. 2 inasmuch as it was given after the other Mixed Relations Test numbered 2 in the Monograph. The results are similar both in time and score.

The average reaction time was 192.4 seconds; the A. D. was 56.25 seconds; the S. D. was 72.38 seconds. There was a time range of 66 seconds to 330 seconds.

The average score was 63.76 per cent.; the A. D. in score was 18.36; the S. D. was 29.46; the median was 65 per cent. The crude mode, or the score that was obtained by the majority was 85 per cent. The scores ranged from 100 per cent. to 15 per cent.

Instead of giving a list of the errors as was done in the previous tests, the different types of erroneous reactions have been analyzed; thus,

1. Repetition of one of the other of the given terms, as, Ear—Hear—Eye—Hear; or Hour—Minute—"Hour."

2. The use of a term that is descriptive of the third word as, H a m m e r—Tool—Dictionary—(Words—E ncyclopedia—Meaning Describes).

3. The introduction of a word that bears a certain relationship to the third word but is not the relation demanded, as, Wash—Face—Sweep—(Clean Broom); or (Sky—Blue Grass—Grows—Ground—In The Park).
4. The insertion of a word which suggests the completion of the thought as in this case.—Buy—Sell Come—(Here—Listen In— Home—Away).
5. There was noticeable also a tendency to insert a rhymed-associate to the third word, as, Straw—Hat—Leather—"Feather": and Hammer—Tool—Dictionary—Tarry: and Little—Less—Much—"Such": and Once—One—T w i c e—"Nice"; Oyster—Shell—Banana—"Anna": and Sky—Blue—Grass—"Glass": and Buy—Sell—Come—"Some."

It is difficult to determine by what process of reasoning this type of reaction was arrived at, inasmuch as the only correct verbal relation that rhymes is the first one i. e. Eye—See—Ear—"Hear" and evidently this first impression "held over." There was, however, no consistency shown by the subjects making these reactions, none of them attempted to find rhymes for all the words supplied.

6. Other failures were due to a lack of linguistic knowledge or an endeavor to define a stimulus word instead of fulfilling the conditions of the test, as, for instance, Little—Less—Much—"Let—Not Enough."

Many of these reactions reveal a groping for the right word, but a failure to grasp the relationship between the first two words clearly enough to reproduce it in an analogous case.

In this type of test the speed of the reaction depends on maintaining the proper adjustment throughout the series. Extraneous ideas must not be permitted to enter. Confusion, therefore, as shown in the failure to "carry over" the crux of the problem, indicates a lack of control and suggests that the subject who cannot form logical connections between ideas is inefficient.

Cancellation of Digits—The figure 2.

In this test the subject is required to cross off the figure "2" in a prepared form which contains this as well as many other digits. The methodology was the customary simple explanation of the problem, and the admonition to "begin at the first line and cross off as rapidly and carefully as possible all the figure "2s." The stop-watch was started when the subject turned over the sheet and stopped when she finished the last line. Some subjects traversed alternate lines from right to left. Two or three attempted to cross off

the digits along the vertical rather than the horizontal lines. This tended to increase instead of lessen the time, involved a waste of energy, and because of a diffusion of attention resulted in many omission of digits. The fact that one's eyes move from left to right in reading makes this the more natural movement in performing this test.

Five subjects complained of eye-strain as the result of the close application required by the task. The time consumed by these subjects did not vary materially from that of the average performance.

The average time, regardless of errors, was 201.5 seconds. The time ranged from 330 seconds to 120 seconds. Eight persons had a perfect score. One subject omitted seventeen twos. The average number of errors was 4.5.

The same system of penalties for errors was employed as that described in the Substitution Test. On the basis of the results of this scoring method, the subjects were arranged in rank order so that comparisons with other test records could be made.

The fact that the more intelligent of the subjects failed to make a high score in either time or accuracy may be due to the rather uninteresting material of the test. The subjects of a lower grade of mentality seized the test with avidity as if saying, as one girl actually did: "Here is something I can do easy." The task would seem to throw some light on whether or not a person is willing to perform routine work.

B. Cancellation of the Letter A.

There were more errors of omission in this than in the previous test. Two persons failed to cross off 32 of the A-s, one omitted 49! The number of errors ranged from one, then, to 49; while thirteen subjects had perfect scores. The average number of errors was nine. The median score on a percentage basis deducting one for every error irrespective of the time required was 90 per cent.; 100 per cent. perfect. The A. D. was 6.95.

The average time scored was 132.5 seconds. There was a range in time of 295 seconds: the best time scored was 68 seconds, the longest time required was 363 seconds. The A. D. in time was 35.19 seconds.

According to the Columbia University records, the A Cancellation Test requires, on an average, 95 seconds. (47).

The test has been given to various groups. Chambers reports the average time for the test as 75 seconds (48); Wallin as 100 seconds (49); while Doll (50), Burt (51), and

Wyatt (52) report the average time for the test as approximately 100 seconds. It is not surprising that the average time for the factory group is so much longer, because, in general, they react more slowly than do the normal members of group accustomed to doing a great deal of writing. The motor co-operations necessary showed among the older women disuse in a marked degree. For them, the task of writing their responses involves an actually laborious application of unaccustomed motor activity.

In reporting the results of this test in his experiments in testing twins (53), Prof. Thorndike did not count errors though they occurred in about one-third of the cases. The time element, he says, is the important factor in the test and accuracy may be made subservient to it.

In scoring the test, for the purposes of this investigation, however, it seemed expedient to use the method of penaltics employed in the previous test. The subjects were given a rank order of merit rating on the basis of the speed and accuracy demonstrated by each individual.

Dr. Weidensall contrasts the performance of her group in this test with that of the Working Girls that Dr. Woolley examined in Cincinnati, and found the Bedford Group are slower in this simple task of checking letters than the working girls were. (31). Only a small per cent. of the Reformatory women were as able as the more efficient type of maid in cancelling letters with accuracy and speed. The scores made by the Factory Girls compared to those of the Bedford Group show marked differences in rate, the factory girls performing the task in a shorter time.

The Bedford Group have a time range of 610.4 seconds to 84.1 seconds which is 4 per cent. longer than the longest time record of the Factory Girls. The best record of the Bedford Group is 1.9 per cent. slower than the best record made by the girls at the Factory.

WOODWORTH-WELLS—DIRECTION TESTS.

The person who is able to carry out directions promptly and accurately is of great value in any business or industrial concern. This firm wished particularly to have a scientific estimate made of each employee in regard to this ability. The three tests arranged by Prof. R. S. Woodworth and F.L. Wells were therefore given. The two Easy Directions Tests were followed by the Hard Directions Test. (54).

The reading time of these tests varied greatly with the subjects, and although as the authors state, "The directions were made as concise as possible in order that the time re-

quired for the mere reading might not be a determining factor," yet in this group there were to be noted individual differences of wide range in the time consumed in reading and grasping the significance of the text. Let us consider each test in turn.

Easy Directions No. 1.

This set of twenty questions was performed correctly by forty per cent. of the subjects. The directions are so simple that there were few mistakes; a failure to react at all was more frequent than a wrong response. Thirty-eight per cent. of the errors were those of omission due either to carelessness or to a lack of comprehension of the problem involved.

Each direction is numbered from 1 to 20. The errors having been tabulated, the questions are arranged in order of their difficulty beginning with that question in which the most errors occurred, which happens, in this case to be the last, No. 20.

Number 20—Cross off the last word in this sentence.

Nineteen per cent. crossed off the last letter in the word Sentence, instead of following directions and crossing off the entire word "Sentence." One omission and a line drawn through the first word rather than the last were the only other errors in this test.

Number 6—Write the sum of these numbers 3—4.

This proved a stumbling block to eleven per cent. of the subjects who wrote one or the other of the following numbers: twelve, one-and-a-quarter, eight, nine, sixty, seventy-five or three-fourths. Nine per cent. omitted the problem.

Number 4. Write —|— over the longer word: It rained yesterday.

Ninety-one per cent. had a perfect score. Six per cent. omitted the question and three per cent. wrote one or the other of these words, "plus," or "over" or yesterday."

Number 3. How many feet makes a yard?

Eighty-four per cent. reacted correctly. Eleven per cent. stated that there are twelve feet in a yard, one said "36" feet and the question was omitted by three per cent.

Number 13. Make two dots between these lines: (lines drawn here).

This was done in 92 per cent. of the cases. Four per cent. made but one dot, instead of two between the lines; while three per cent. put two dots outside of the lines and one per cent. did not do anything at all.

Number 12. Write g on the egg-shaped figure (three designs given.)

Ninety-four per cent. did this; three per cent. wrote "g" on the wrong figure, and the remaining 3 per cent. did nothing.

There were no errors in the other directions. The average score was 89.14 per cent.; the Median score was 90 per cent.; the A. D. was 3.84 and the S. D. was 7.93.

The average reaction time was 92.4 seconds; there was a time range of 55 seconds to 296.2 seconds. The A. D. was 27.6; the S. D. was 35.34.

The average reaction time required by the Woodworth-Wells group of educated adults was 72 seconds, with limiting records of 46 and 114 seconds.

Easy Directions Test No. 2.
(Beginning "Cross out the smallest dot").

This test is of practically the same degree of difficulty as the preceding test.

As in the other test, errors of omission were more frequent than any other sort.

The directions are numbered from 1 to 20 inclusive. Those in which errors occurred are noted below in the order of their difficulty, beginning with that one in which the most errors are found.

14. Write O after the largest number: 3, 86, 12.

Thirty per cent. wrote O above 86, ten per cent. wrote it below. Two per cent. wrote O after the 3, and two per cent. crossed off the number 86, while two per cent. failed to react at all.

15. Mark the name of a large city: New York, painter.

Instead of doing this the very easiest and most obvious thing, twelve per cent. wrote one or the other of the following words: New York City, Paris Fashions, Pittsburg, England, Chicago.

7. Write any word of three letters.

Four per cent. wrote words of four letters such as White, Word, Good and Four. Others wrote correctly a variety of words of three letters,—the words chosen were: eat, two mar, eye, the, how, are, got, his, dot, any and cat. The last word was the most popular.

There were one or two errors in the remaining directions, but they are not significant nor indicative of the intelligence of the group.

The average score obtained by this group was 85.45 per cent. The median score was 83.72 per cent.; the crude mode was 100 per cent. The A. D. was 6.84 and the S. D. was 12.96.

The average time for the group was 161.9 seconds; the A. D. was 44.85 seconds, the S. D. was 56.78 seconds. The time range was from 44" to 235".

The average time for the Bedford Group was 151.9 seconds, with limiting records of 55 to 397.4 seconds. (31).

When the extreme simplicity of both this and the previous Directions Test is considered, the difficulty which these girls experience in handling written work is apparent. In observing the reactions of the various members of the group, the reading over of the directions seemed to consume the major part of the time. Dr. Weidensall states that the best quarter of her subjects at Bedford read over the text in 40", the median subject in 50", and the poorest quarter in 90". The factory subjects showed a tendency to read and reread the instructions, and although admonished to "work as rapidly as possible," this inability to grasp quickly the meaning of the printed words seemed to be the cause for the retarded and prolonged reaction times.

3. Hard Directions Test.

This test measures the ability to read complex sentences and understand them and appreciate, when alternate clauses are given, that a choice is to be made. It is more important that the idea of making a choice be comprehended than that the correct response be written. In scoring a system of allowing half credits was arranged so that the significance of this feature of the test might be evaluated properly.

An analysis of the errors made in the test follows. The "directions" have been numbered and are arranged in the order of their difficulty, those in which the most mistakes occurred placed in gradation from a direction in which 51 per cent. of the girls failed to one with only 8 per cent. failures.

10. "Write NO if 2 times 5 equals 10," was an injunction disobeyed by 51 per cent. of the subjects all of whom wrote "yes."

8. "Give the wrong answer to this question, How many days are there in the week?" elicited a response of "7" from

48 per cent. There were five omissions and the remainder of the subjects wrote either a 6, 9, 5, 10, 8 or 2. The digits 6 and 9 were the favorites. It is to be noted that no one wrote 1, 3 or 4.

20. "Write the first letter of your first name and the last letter of your last name," proved extremely confusing. Eleven per cent. failed because they wrote either their whole name or else omitted writing anything at all. Thirty-five per cent. wrote the initial letters of both their first and last name and could, therefore, receive but half-credit.

14. and 15. These two directions, the one dependent upon the other can best be discussed together: "Notice these two numbers: 3; 5. If iron is heavier than water, write the larger number here.....,"but if iron is lighter write the smaller number here.....

A subject, who recognizing the force of the alternate questions, as, for instance, a subject who indicates that water is heavier than iron because steamers are made of iron and float, should receive more credit than one who makes no choice. Some subjects placed numbers in both of the blank spaces, some put one or the other number in either space. Since this type of response showed a total lack of comprehension of the idea of choice both questions were called failures. One-third of the girls were unable to grasp the idea of choice and failed in both, and three per cent. omitted to fill in either space.

17. and 18. "Give the correct answer to this question: Does water run uphill?....and repeat answer here....." Twenty-five per cent. either omitted any reply or else wrote the word "down-hill" in the first blank only, and did not repeat the aswer.

19. "Do nothing here (5 plus 7 equals....), unless you skipped the preceding question." This question came next in order of difficulty, with a twenty-one per cent. of failures. Five and seven were added, or else, ignoring the significance of the plus sign, the digits were one or the other of them copied in the space, or else the plus sign was interpreted as a sign of multiplification, and "35" appeared in the blank space.

11 and 12. "Now, if Tuesday comes after Monday, make two crosses here.....(11).; but if not, make circle here.... or else a square here...(12)." These alternate choices caused a good deal of confusion. Twenty per cent. failed to see that a choice must be made, and thoughtlessly, made the crosses and inserted a circle or a square or both in the blank spaces. Nine per cent. omitted the question entirely.

16. "Show by a cross when the nights are longer: in summer?.....or in winter?".....The popular error in this reply was to put a cross in both blanks, which was a com-

plete failure for 19 per cent. Twelve per cent. indicated that nights are longer in summer than in winter.

3. "Then, if Christmas comes in March, make a cross right here.....but if not, pass along to the next question, and tell where the sun rises.....(4)." Twelve per cent. made a cross in the first blank and also wrote in the second blank one or the other of these phrases: "in the sky," "in the morning," "in the north," "south," or "west"; six per cent. omitted to answer either part of direction, while 16 per cent., recognizing the idea of choice to be made, received half-credit.

7. "Write YES, no matter whether China is in Africa or not....." Sixteen per cent. wrote No, while 3 per cent. failed to write anything at all.

5 and 6. "If you believe that Edison discovered America, cross out what you just wrote (5), but if it was some one else, put in a number to complete this sentence: "A horse has ....feet." (6). No one crossed off what she had just written, but five per cent. did cross off the word "Edison," and in its stead wrote "Columbus." There were several interpolations of a number between the semicolon and the sentence, "A horse etc." The horse was given but two feet in 5 per cent. of the cases; however, there were only two omissions and hence 90 per cent. of the subjects reacted correctly to both 5 and 6.

9. Write any letter except g after this comma." This seems a very simple and straight-forward direction yet there were 6 per cent. of the subjects who proceeded to write g in spite of the injunction to do otherwise. Four per cent. wrote nothing.

2. "Put a comma after the longest of these three words: boy mother girl." There were types of failure: a comma was placed above the longest word, or after each one of the words, or after the word "girl," or nothing at all was done. The total number of failures amounted to 9 per cent.

1. "With your pencil make a dot over any one of these letters: F G H I J." This direction, proved the easiest to interpret, met with failure in 8 per cent. of the cases. Dots were placed over each one of the letters, while one girl dotted most of the given letters as they appeared in the text. There were no omissions.

The average score in this test was 77 per cent.; the A. D. was 9.41; the S. D. was 11.32; the median score was 80 per cent.; the scores ranged from 100 per cent. to 45 per cent.

The average time score was 262.12 seconds; the A. D. in time was 48. 58 seconds; the S. D. in time was 53.49";

there was a time range of 70 seconds to 719.5 seconds.

Woodworth and Wells quote 107 seconds as the average time for their group of college students. We can report but three subjects with as good a time score as this average made by the students, and Dr. Weidensall had only two subjects among the Subjects examined who require as short a time. Of the Reformatory women all but one were slower than 134 seconds which was the poorest time-score of the college students. (31). The average time for a group of sixty normal students at the Chicago Normal College is 118.8 seconds (31) which is slower than the Woodworth and Wells subjects, but a better time-record than was made by the majority of our subjects, since only the three mentioned above performed the test in as short a time. The index between time and accuracy derived from this test shows that the highest score was not obtained in the shortest time, nor on the other hand was the slowest person the most accurate. The relation between time and accuracy is discussed under a separate heading.

The Bedford Reformatory Group performed the test in 274.85 seconds, average time for the group, which is 12.63 seconds slower than the average time required by the Factory Girls. (31).

LINGUISTIC TESTS.

A. The Trabue Completion Test.

The Trabue Completion Test beginning "The sky..... blue." appeared to be an interesting form of amusement to most of the subjects, for they went smilingly through the test, making errors often but generally writing something in the blank spaces.

Since the English spoken by the majority of the Factory workers contains much incorrect phraseology, it seemed wise to overlook minor details in the results, and count as errors only impossible and meaningless interpolations. It was common to find the plural subject of a sentence completed by a verb in the singular, and vice versa. A logical connection of ideas rather than grammatical rectitude seems a fairer criterion than to lay undue stress upon linguistic construction. This group made some interpolations that are unique, and as they are not noted by other investigators, Dr. Trabue was consulted in regard to their proper valuation. In general, however, the prescribed method was utilized. A system of half-credits was found necessary.

The sentences vary so much in point of difficulty that they must be discussed individually.

1. The sky.....blue.

No one failed to interpolate a word that made sense and "looks", "was" and "is" were used with about equal frequency.

2. We are going....school.

The only errors in this sentence were the insertion of "home" and "play" by 3 per cent. of the subjects.

6. The.....is barking at the cat.

Every one had this correct.

8. The stars and the.....will shine to-night.

Four per cent. of the subjects inserted the word "sun" 21 per cent. supplied the word "stripes", while "moon" was written by the remaining 75 per cent. of the subjects.

10. The.....sings a song every morning.

The only error here was the use of the word "will".

12. Good boys.....kind to their sisters.

Here one finds "be", "run errands", "and so are" inserted instead of the very obvious "are" which was given correctly in 95 per cent. of the cases.

14. When the.....grows older he.....be a man.

Four per cent. inserted a proper name in the first blank over-looking the significance of the definite article "the," otherwise there were grammatical errors only in this and the two succeeding sentences.

16. The boy will.....his hand if.....plays with fire.

The word "wash" was used in 8 per cent. of the cases.

18. The best.....to sleep is at night.

This was correct in 78 per cent. of the reactions, the remaining 22 per cent. inserted the word "thing", or "cure", "his", "is", "to go", "place", "way", or "hour", showing an inability to comprehend the meaning of the sentence.

20. The little.....and his dog.....running a race.

The only errors in this sentence were those in which a plural subject was followed by a singular verb and vice versa.

22. Time.....often more valuable.....than money.

Fifty-two per cent. had these two insertions right; the rest of the subjects wrote one or the other of these words "Work," "Come," "Has," or "Are" in the first space and "in" or "to" in the second space quite os often as "Than."

24. The.....rises in the morning and.....night.

The words "Boy", "Girl" or "Bird" were supplied by 40 per cent. of the subjects "rising" in the morning and "sleeping" at night. The majority, however, wrote "Sun" correctly, while 50 per cent. inserted "sets". The remaining 50 per cent. wrote "lowers" or "moon" in the second blank.

26. Boys who play.....mud get their hands.....

These blanks were filled in correctly by 50 per cent. of the girls. There were various insertions descriptive of "mud", such as "nasty," "street," "black," "much," "bad," "soiled"

and "yard". In the second blank "bad", "full" and "soiled" were chosen in about equal proportions.

28. The little....had....nothing to....; he is hungry.

But 6 per cent. failed to recognize the significance of the pronoun "he" and wrote "girl" or "creatures" in the first blank.

30. The boy who.....hard.....,do well.

This sentence was confusing to 70 per cent. of the subjects who inserted "have", "is", "tries", "has",, "studies", "eats" or "thinks" in the first blank, and "always" "work" or "does" in the second.

32. One's.....do.....always express his thoughts.

Eleven per cent. wrote the word "who" in the first space and fifteen per cent. the word "mind". Seventy-one per cent. supplied meaningful words, but the remaining 18 per cent inserted either "ideas," "opinion," "self," "actions," "speech" or face," "Not" was written invariably in the second blank space.

34. It is a.....task to be kind to every beggar..... for money.

Words that made sense were inserted by 33 per cent., while 50 per cent. wrote words that could be given half-credit. The remainder left the spaces blank.

36. Worry.....never improved a situation but has .....made conditions.

Thirty-five per cent. of the subjects completed this sentence correctly; twenty-five per cent. received half-credit and the remaining thirty-seven per cent. failed to insert any words that could be accepted; ten per cent. made no attempt to supply any words.

The average score was 80.33 per cent; the Median Score was 84 per cent. The scores ranged from 100 per cent. to 34.8 per cent. The A. D. in score was 13.76 per cent. The S. D. was 24.62 per cent.

The average reaction time was 464.5 seconds; the median time rate was 340.5 seconds; the time scores ranged from 293 seconds to 787 seconds.

LINGUISTIC TESTS—ENGLISH-AFRICAN PROVERB TEST.

No published data of this test arranged by Dr. Ruger of Columbia University has appeared. Prof. Thorndike included the test in a set given at Teachers' College, but no detailed report of the results of the test are noted. (55).

The Experimenter beside presenting it to this group of Factory girls, gave the test to a group of persons employed at the Willard Parker Hospital. The reactions of these two

groups may well be compared.

The directions were read to the subjects, and each one was asked in turn if she understood what was to be done. If she did not comprehend the problem, she was asked to read aloud the first of the English Proverbs, and then find among the African Proverbs the one that seemed to mean most nearly the same thing. Stimulated by a suggestion as to a method of procedure, the subjects invariably did something with the test.

The Factory subjects were asked after they had read over the test if there were any words which were unfamiliar. Six girls announced that they did not know "enchantment," and "Leisure" and "Folly" were new words to three girls. It is possible that other persons failed also to grasp the meaning of all the words but were too shy to admit their ignorance. It is obvious from an inspection of the accompanying table that some of the confusions occurred through a lack of comprehension of the phraseology.

Results of the Test administered to the Factory Girls are as follows:

5 per cent. had a perfect score.
5 per cent. had eleven of the analogies right.
1 per cent. had ten of the analogies correct.
3.33 per cent. had nine of the analogies correct.
1 per cent. had eight of the analogies correct.
5 per cent. had seven of the analogies correct.
8.33 per cent. had six of the analogies correct.
31 per cent. had five of the analogies correct.
8 per cent. had four of the analogies correct.
21 per cent. had three of the analogies correct.
10 per cent. had two of the analogies correct.
15 per cent. had one of the analogies correct.
5 per cent. had none of the analogies correct.

Each pair may now be discussed in the order in which the English Proverbs occur in the text.

A. Married in haste we repent at leisure.

11. Quick loving a woman means quick not loving a woman.

These two seemed by far the easiest proverbs to analyze and 87 per cent. paired them correctly; only 6 per cent. omitted A entirely, 3 per cent. paired 11 with "First catch your hare" (G), and 3 per cent. paired 11 with "Distance lends enchantment to the view" (K). There is a semblance of humor in the latter combination, unless one chooses to ascribe it to chance! Every one had heard the English version of the proverb and the expression offered no complication to most of the subjects. Except to the very stupid, the African interpretation of a hasty marriage was perfectly obvious.

B. Answer a fool according to his folly.

6. If the boy says he wants to tie the water with a string, ask him whether he means the water in the pot or the water in the lagoon.

This analogy was given by 34 per cent. of the subjects. Number 6 was paired with "Distance lends" etc. (K) by 9 per cent; and was coupled as well with A., C., G., H. and J., respectively, by 8 per cent. of the subjects; while 3 per cent. paired No. 6 with L. and 3 per cent. with M., and 4 per cent. coupled it with D.—"Out of the frying pan" etc. Practically one half of the subjects failed to make any combination with No. 6 and 8 per cent. made no attempt to dispose of B.

C. One swallow does not make a summer.

1. One tree does not make a forest.

Fifty-one per cent. had these two paired correctly. Twice No. 1 was combined with "Birds of a feather etc.," because as these subjects explained, both proverbs contain the word "Bird." Eighteen per cent. failed to pair C. with any other proverb, and twenty per cent. failed to find an analogy for Number 1.

D. Out of the frying pan into the fire.

8. He runs away from the sword and hides himself in the scabbard.

Only twenty-three per cent. had this correspondence correct. Fourteen per cent. combined E. and 8. Six per cent. thought J. and 8 analogous. Three per cent. paired 8 with I., and three per cent. paired it with M. and three per cent. with K. Twenty-one per cent. failed to make any combination for D., and twenty-nine per cent. omitted No. 8.

E. Robbing Peter to pay Paul.

13. No one should draw water from the spring in order to supply the river.

These two proverbs were arranged correctly by twenty-eight per cent. of the subjects. Ten per cent. paired M. with 13; and several paired it quite at random, apparently. J. was combined with 13 by 6 per cent. of the subjects, and D was combined with 13 by four per cent. Twenty-five per cent. omitted E. entirely, and thirty-one per cent. did nothing with 13.

F. Birds of a feather flock together.

9. A fool of Ika and an idiot of Lluka meet together to make friends.

Thirty-one per cent. had this analogy correct. Twenty-three per cent. paired 9 with B., "Because," said they, "both contain the word 'fool'; six per cent. paired 9 with I., arguing that for a fool and an idiot to meet is "adding insult to injury." Three per cent. joined 9 with A., "For," said these subjects, "persons marrying in haste" are a fool and an idiot meeting. Number 9 was combined with E. an equal number of times as "left over" propositions that could not be disposed of other-

wise. Fifteen per cent. failed to pair F. and seventy per cent. omitted Number 9.

G. First catch your hare.
2. I nearly killed the bird. No one can eat nearly in a stew.

This pair of proverbs proved more difficult to analyze than did any of the others. Number 2 was paired at least once with every other proverb, and with F. in seven per cent. of the cases. Both F. and E. contain the word "Bird" which was given as the justification for this arrangement. One person insisted that "No one can eat nearly in a stew" means that no one can eat when nearly in a fret or hurry—i. e. in a stew. Eighteen per cent. made the right combinations. Forty-two per cent. did not pair G. at all, and twenty- six per cent. left out Number 2.

H. Sour grapes.
7. Cocoanut is not good for bird to eat.

This pair appeared as difficult to analyze as the previous one inasmuch at the same percentage, (thirty-one), arranged this and the former pair correctly, although the combinations made were not as varied as in the C. and number 2 Proverbs. "Sour grapes" proved a fairly familiar saying, but "Cocoanut" suggested the idea of food, "Milk for babes" (L) to twenty-one per cent. of the subjects, and to six per cent. the word "Bird" made the combination of Number 7 with F. seem the logical one. Six per cent. of the subjects coupled Number 7 with E., G. and J. respectively. Eighteen per cent. made no disposal of H. and thirty-one per cent. none of Number 7.

1. Adding insult to injury.
10. The ground pig said: I do not feel so angry with the man who killed me as with the man who threw me to the ground afterwards.

Twenty-one per cent. found this analogy. Twelve per cent. paired Number 10 with J., and seven per cent. paired it with M., and seven per cent. with G. The last two proverbs refer to animals, the subjects who made these combinations remarked, so why not put them in pairs! Six per cent. coupled Number 10 with D., and three per cent. with B., and H. Fourteen per cent. failed to dispose of I. in any fashion and fifteen per cent. did nothing with Number 10.

J. Curses come home to roost.
3. Ashes fly in the face of him who throws them.

Only twelve per cent. of the subjects combined these two. Fifteen per cent. paired Number 5 with D., because the word "ashes" suggested "fire" and there was none other that seemed to "fit as well." Each one of the English Proverbs was paired with Number 5 at least once, and, as one girl expressed it, "I put down all I was sure were right, and the rest I just settled any way at all! This method of elimination was

quite the usual one. Thirty-seven per cent. left J. a blank and thirty-one per cent. omitted Number 5.

K. Distance lends enchantment to the view.

4. Distance firewood is good firewood.

Upon the supposition, perhaps, that proverbs containing the same word, or one of its derivatives, should be paired, twenty-five per cent, of the subjects combined these two immediately. Fourteen per cent. followed a similar line of reasoning, allied Number 4 with D., since "both have to do with fire." An equal number of persons thought Number 4 and F. were analogous, while in three per cent. of the cases Number 4 was paired with B., C., H., and M. Thirty-four per cent. omitted K entirely, and fifteen per cent. failed to pair Number 4.

L. Milk for babes.

12. If the stomach is not strong do not eat cockroaches.

These two proverbs are too abstruse, apparently, for this group since but ten per cent. grasped the analogous meanings. "Milk for babes" was paired with Number 3. in twenty-six per cent. of the cases, an association of ideas that can be understood. Twelve per cent. made no disposal of L., and thirty-six per cent. none of Number 12. Twelve per cent. combined Number 12 with H., while six per cent., recalling the pangs of indigestion, paired Number 12 with "Curses come home to roost" (J.), or as one girl expressed it, "If you eat what does not agree with you, you always have to pay!" Six per cent. reasoned that giving the "stomach cockroaches to eat" was analogous to "adding insult to injury" (I). There were other combinations that the introspections showed were studied, but the majority were chance associations only.

M. We can all endure the misfortunes of others.

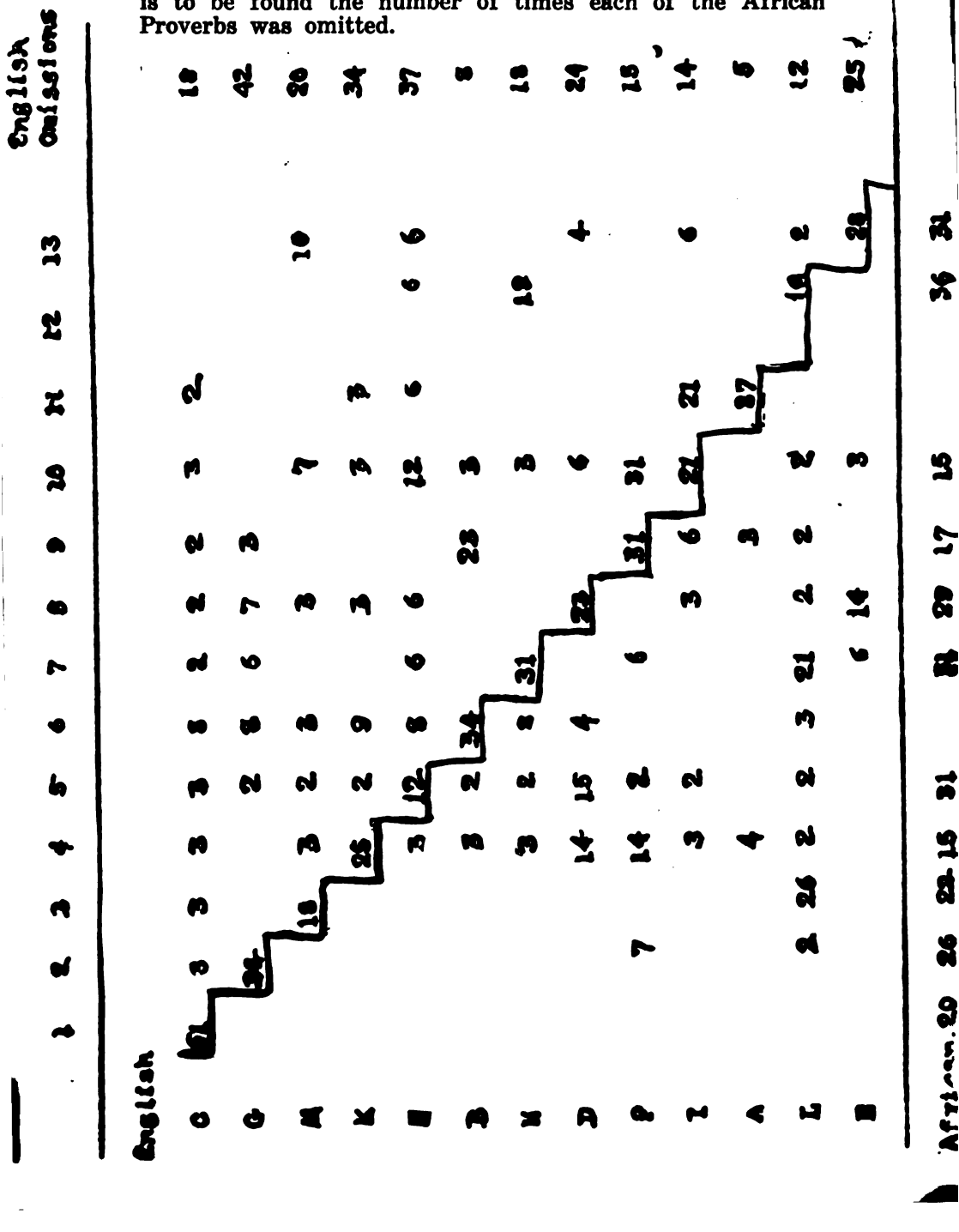
3. Full-belly child says to hungry-belly child, "Keep good cheer."

Eighteen per cent. paired these two proverbs, but twenty-six per cent. combined Number 3. with L., and M. was coupled with every other proverb in the African Group, except Numbers 1 and 11, at least twice. Twenty per cent. failed to make any combination at all with M., and twenty-two per cent. failed to connect Number 3 with any other proverb.

Arranged in the order of difficulty, beginning with the proverb that was compared correctly most often, the English Proverbs appear thus:—A., C.,; B., H. and F.; I., K., and D.; G. and M.; J. and L.

The table given below reveals in a graphic manner the number of correct analogies and the omissions that were made, as well as the confusions and errors that occurred. These have been noted, but the tabulation gives a better idea of comparative difficulties in analysis than can be derived from a purely verbal elucidation. The zig-zag line through the center

designates the correct pairing of the proverbs. In the vertical column to the right, the number of times each of the English Proverb was omitted is given, and, at the bottom of the table, is to be found the number of times each of the African Proverbs was omitted.



It was noted that the first six persons who tried the test completed it in less than 13 minutes, the next six persons made neither addition nor correction after a period of thirteen minutes had elapsed so a time-limit of thirteen minutes was set. In the experiment conducted under the direction of Prof. E. L. Thorndike (55) an allowance of twenty minutes was made for this test in order that the measurement might be wholly of the quality of the work done irrespective of the time element. The Factory situation did not warrant the devotion of so much time to this one test. The Experimenter feels confident that the test-results would be practically the same had more time been allowed the factory girls. No time limit was fixed for the Hospital Group and their time ranged from 900 seconds to 185 seconds, with an average time record of 543.14 seconds; the A. D. was 153.39 seconds.

The time consumed in the execution of the test by the Factory Girls ranged from 780 seconds to 312 seconds. The average time for the test was 581.149 seconds; the A., D. was 153.39 seconds.

The test was scored on the basis of 100 per cent. perfect, or an allowance of 7.7 per cent. each correct answer.

The average score for the Factory Group was 29.5 per cent.; the A. D. was 6.45.

The average score for the Hospital Group was 72.94 per cent.; A. D. 18.93; the S. D. 26.54.

VI. DESCRIPTION OF THE SPECIAL TESTS.

1. Judgment Test.

The Foreman in the Selecting Department had observed considerable variation in the ability of the girls under his direction and wished to determine how they compared in efficiency as measured by a series of special tests, and whether any other girls possessed aptitude for the work. In will be recalled that girls in the Selecting Department are obliged to sort feathers according to their length and width, regardless of their quality. This is the most important work done in the department, since an error of a quarter of an inch in the estimation of the dimensions of a feather means quite a loss of money to the firm, hence to assist the inexperienced worker there is a foot-rule scale on each girl's table. She is expected to lay the feather on the scale if she cannot make an independent decision. The more exact and expeditious the sorter becomes the greater is her individual value to the firm, and the truly efficient sorter forgets that the scale exists. Now, although practice in this type of work is of the utmost importance, the aptitude for making quick decisions varies with the subject.

In order to determine the degree of variation in forming such judgments as are demanded in this department, the following test was arranged.

Twenty strips of fairly heavy card-board cut to approximate the sizes of the feathers examined during the ordinary routine of the day were prepared. (See measurements given in the Appendix.) These strips were numbered to correspond with the scale in actual use in the Shipping Department, so that they might be readily identified by the Examiner and checked up quickly.

Each strip of card-board was held at distance of eighteen inches from the subject and she was asked to estimate the length and width of the strip, writing her appraisement on the specially prepared sheet that had been supplied her for this purpose. The strips were exposed for ten seconds. Although rapidity of decision counted in favor of the person examined it did so only when quickness of answer coincided with exactness of answer. At no time was wrong answer rapidly arrived at counted as anything but a failure against the subject giving it.

In scoring, 5 per cent. credit was given for each correct judgment, and for every error of half-an-inch in an estimation a deduction of one per cent. was made from the score.

Sixty-seven per cent. of the judgments of width were correct and fifty-two per cent. of the judgments of length.

The average error in judgment was 1.53 inches, with a range in error of 1.88 inches, or errors that ranged from .62 inches to 2.5 inches.

The average over-estimation of length was 1.83 inches, while the average under-estimation was 1.09 inches. The range of error in under-estimating the length of the strips was from .45 inches to 3.04 inches. The average error in over-estimating the length was from .63 inches to 2.53 inches.

In the matter of under-estimating widths there was an average error of .87 inches, and in over-estimating widths of .869 inches. The range of error in over-estimations of width was from .58 inches to 1.68 inches. All of which goes to show that there was a somewhat greater tendency to over-estimate than under-estimate the dimensions.

Ritter (56) reports in an experimental study that there is a definite inclination to over-estimate vertical distances and under-estimate horizontal distances. This may be due to the increased effort required in moving the eyes up and down instead of in the more accustomed left to right movements used in reading. The differences in judging the length and width of the feathers made by these girls are indicated in this table:

	Over-Est. of length	Under-Est. of length	Over-Est. of width	Under-Est. of width
Average	1.876 in.	1.09 in.	.869 in.	.777 in.
A. M. D.....	.611 in.	.534 in.	.188 in.	.428 in.

The time required for the entire test averaged five minutes and three seconds. The quickest person made her decisions in three minutes and three seconds and the slowest in seven minutes and twenty seconds.

The index between time and score was higher in this than in the purely intellectual tests, showing a correlation of .63 (positive), with a Probable Error of plus or minus .059.

It is possible as a result of the analysis of the records made in this and the two following tests to discriminate with much exactitude regarding the output of the individual worker.

As has been observed, the ability to form quick and accurate judgments in the matter of color, form and quality is a most valuable asset in a factory employee. The table showing the individual correlations of the girls and their rating by the Firm illustrates the practical application of this particular group of tests.

It was observed that those persons who hesitated over their decisions were the ones who made the greatest number of errors; but, on the other hand, those who passed judgments with scarcely a glance made more errors than the subject in the Median Time Group who are neither slow, nor yet too hasty.

There are a preponderating number of over-estimations of both length and width, showing what might be interpreted as a tendency toward exaggeration. The Manager, noting this tendency when inspecting the results, stated that he wished there were tests to reveal the presence of this trait, for among the "outside men" employed by the Firm unnecessary difficult arose because either they over-estimated the promptness with which orders could be executed, or else they exaggerated the quality and character of the goods.

II. Sorting Test.

A hundred strips of fairly heavy card-board were cut to correspond with twenty of the most frequently used feathers such as were described in the previous test. There were, therefore, five strips of each size. These strips were laid on the table in front of the subject in a heterogeneous mass just as the feathers are received from the stock-room. Each girl was directed to sort the strips according to size, treating them as though they were feathers and laying those of like dimensions in the same pile. They were told to work as quickly and carefully as possible.

This test requires the exercise of the same sort of judgment, concentration and motivation that is involved in the performance of such a piece of factory activity as this: suppose that an order for a hundred feather fans has been received, and several thousand feathers are sent down from the stock-room each sorter is given a share and instructed to lay feathers of the same dimensions in a pile, the work must be done expeditiously, and it is important that the matching of the feathers be perfect, inasmuch as the finished product depends on the symmentry of the feathers selected. It is important that the sorter possess excellent discrimination and the capacity to make swift decisions. That this test shows individual variations in ability to perform work of this order was proved conclusively by the correlations obtained.

The average time for the test was 255.33 seconds, with an average deviation of 80.63 seconds. There was a range in time of from 70 seconds to 360 seconds.

The sorting was on the basis of 100 per cent. for five neat piles of cards of similar sizes, irrespective of the time consumed in the sorting process.

No. one made less than five piles, but 35 per cent. made more than five. A system of penalties for extra piles was used, and each person making more than five piles had an additional five seconds for each extra pile added to her time-score.

The girls were arranged in order of merit on the basis of their standing on the Time-record.

The moderate workers did the best sorting, the swiftest made errors and their piles were untidy and slip-shod; the very deliberate, slow, painstaking workers saw differences where none existed.

III. Card Sorting.

A simpler test, that of arranging two packs of ordinary playing cards in suits was given to each one of the factory group.

Skill in this test depends much on practice but such psychological facts as neatness, system, or the reverse, may be noted, and the observations of this character made by the Experimenter in individual cases aroused the approval of the Foremen.

The clumsy girl let her cards fall frequently, or fumbled them. A lack of decision and keen perception caused loss of time and most of the errors. The sorting of cards is a simple eye-hand co-ordination exercise very like the work performed by the girls in the Selecting Department. The amount of discrimination necessary is reduced almost to a minimum, and in this regard it is more like the sorting of the Raw

Feathers than it is like the sorting of the finished products, the dyed and curled feathers.

The time ranged from 55 seconds to 181 seconds. The errors ranged from ten to one. By applying the Method of Penalties (46) described by Dr. Munsterberg fourteen seconds was added to each score expressed in seconds for every single error made. Upon the basis of the time-scores, the subjects were arranged in rank order of merit. The test was then correlated with the Criterion and the other tests.

This test helps to differentiate between the quick, neat, careful and resolute worker and the person who is vacillating, slip-shod and slow. It is distinctly an index of manual dexterity. The relation between Time and Score in this test shows an unusually high index, .75 P. E. .0397.

If an individual girl is being tested in order to determine into which of the Departments of the Factory she will fit best, this test is most illuminating. It throws light upon the girl's temperament and her type of performance reveals whether she will do good work as a Selector or be better employed in some other department of the factory.

The median score in this test is represented by the time score, 134.46 seconds. Unless a girl performs the test within this time limit, she will not excell as a selector.

IV. Discrimination of Color Test.

The ostrich feathers are received in their natural condition, and the work of dyeing, curling and dressing is all done at the factory.

After the first sorting of the raw material, the feathers are sent to the Dyeing Department where they are colored and then steamed. When these processes are finished, the feathers are again sorted and arranged so that they may be fashioned into the designs ordered.

There is sometimes a very slight difference in shade that is rather difficult to detect. An expert sorter must be able to differentiate between the slightest variations in color as well as in form and size.

In order to test the ability of the Selectors to do this, a Discrimination of Color Test was devised. Pieces of knitting yarn of various colors, fifteen inches long were tied in small bows. There were four sets of ten shades each of Blue, Purple, Orange and Green, which were mixed with seventy-five bits of yarn of other colors. Each set contained exactly the same number of pieces of yarn, namely, eighty-five small bows of yarn. These were placed in an envelope on the outside of each of which was written either the word ORANGE, PURPLE, BLUE or GREEN. In the envelope bearing one or the other of these color-names all the shades of the particular color specified were placed.

The directions were given orally thus: Each one of you has been given a sheet of paper. Write your name on the paper. You have been given an envelope also that contains bits of yarn of different colors. On each envelope is written the word ORANGE, GREEN, PURPLE or BLUE. That means that in each one of the envelopes there is a set of various shades of one or the other of these colors. Those of you who have an envelope marked "PURPLE" are to put all the shades of purple in a pile on the sheet of paper. Those of you who have an envelope marked "Blue" all the shades of blue, and so on in the same way with the "Orange" and the "Green." Pay attention to no other color than the one you are told to select.

Do you all understand? When I say "Ready," empty the contents of the envelope on your table and begin at once to sort your color. Remember to put the yarn you select on the white paper. When I say "Stop," do so at once. Sort the colors as quickly and as carefully as you can.

One minute was allowed and in that time one person had sorted all the colors. The other girls had selected from 42 per cent. to 93 per cent. of the colors.

Since it was desirable to ascertain those persons who were capable of making quick decisions coupled with a fair degree of accuracy, special stress was placed upon the time-element when the directions were given.

After a girl had sorted one set of colors she was given another set until each girl had had each set of colors to sort. This was done to see whether there was more difficulty in distinguishing between the shades of one color than another, and whether there was a gain in time through practice. The shades of purple seemed more easily and quickly selected inasmuch as 85 per cent. were picked out on an average by the group in the allotted sixty seconds. The other colors in their apparent order of difficulty were green, orange and blue, with an average of 80 per cent.; 73 per cent.; and 60 per cent. selected within the time-limit.

When the results of the four color sortings were averaged the subjects were arranged in rank order according to the record made by each individual in the entire set of colors.

This ranking was then compared by the rating given to these girls by the members of the Firm.

V. Motor Control Test.

Most of the work in the factory is monotonous. The same type of movement is repeated over and over. The ability to stick to the task in spite of its extreme monotony, and create interest where none is apparent, is possessed by

some persons but is totally lacking in others. A girl, who because of a nervous temperament is unable to keep repeating the same form of activity, cannot do good work as a Bender or Finisher of Feathers; neither can she do her best work in the Selecting or in the Boa Department for as our Job-Analysis indicated, practically all of the workers, save those who run errands or are in the Fancy Department, must do the same thing over and over again.

In order to determine which of the employees are best suited for steady and monotonous labor, a new and very simple test was presented. On a sheet of 8 x 11 paper, 1088 oblique lines one-quarter of an inch long, of this character / / / / / were arranged in 29 rows, thirty-eight in a row. The test was described to the subject, the sheet given to her face down, and she was directed to make a cross out of each oblique line as quickly as she could. This was given to small groups though it may well be used as an individual test.

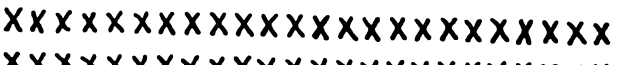
The subjects were told that the Experimenter would announce the thirty-second periods, and they were to check at what-ever point they happened to be working when the number was called. In this way, comparisons between subjects could be calculated, the effect of practice noted, and the individual differences in procedure observed. The length of time allowed for the test in optional with the experimenter. Five periods of 30 seconds each, or two minutes and a half was the allotted time for standardization in this case, although certain subjects were permitted to finish the sheet.

There was not as great a variation in the amount of work accomplished in two and a half minutes as there was in the quality of the work performed. (Five grades of achievement were arranged. See Page 57.

The medians for these subjects averaged 294.875 strokes.

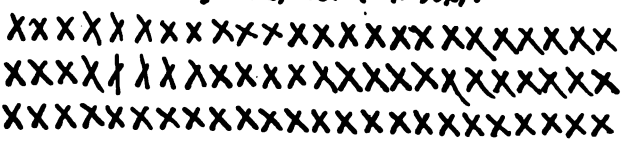
The general appearance of the sheet proved to be the most enlightening and important feature of the test.

Grade I. The most excellent type. Steady, even cross-lines maintained throughout. (90% and over).



Similar to execution of Subjects 166; 138; 147 etc

Grade II. Even and Steady in spots. (80%-90%).



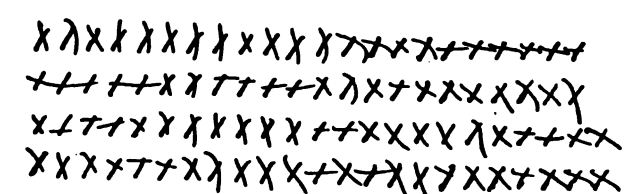
Grade III. More variability than Grade II. (65%-80%).

Execution 112; 173; 106; 107; 127.

Grade IV. Still greater degree of variability. (30%-65%).

Similar to execution of 158; 157; 142.

Grade V. The poorest execution. (Below 30%).



From the records of 119; 143; 151.

The papers were graded by three judges, foremen and directors in the factory. First, they graded the papers on the basis of neatness; Second, according to the number of strokes made; Third, according to the length of the stroke, the shorter and more firmly the stroke was made, the higher was the grade given to the paper; Fourth, the uniformity of the cross stroke, when the strokes were approximately the same length above and below the oblique and when they crossed at about the center of the oblique they were rated higher than when the reverse was true. The Judges were asked to arrange the papers in rank order of merit and were asked to consider first one characteristic then the other in the different characteristics mentioned above. They were requested to pay no heed to any attribute other than the one specifically designated.

After each grading, the Experimenter noted on the back of the sheet the rank order given by each judge in each of the features postillated. The different ratings were then averaged, and a complete scale derived. This ranking was compared with the results obtained by the same employees in the tests as well as with the rank order on the Criterion Scale.

Personal traits of character are revealed by even a quick inspection of these test sheets, and a capacity for precision of movement can be noted as well as the subject's attitude toward monotony. Many of the girls began the test in a regular and steady manner but grew careless after a few lines were done and continued the task in a slip-shod fashion. Steadiness of motor control in these persons was so limited that the monotonous work of continually handling feathers palled on them and as the indices of correlation which are discussed, shows, this test is the best index of capacity to perform this particular grade of factory work that has been found.

The test met with the highest verbal commendation from the firm. Irrespective of any mathematical calculations in connection with this test, the Steady Workers who tended to make rather even, well-balanced lines and the careless, erratic workers who were inclined to change their methods of crossing the lines perhaps several times during the test, could be sorted out after a brief inspection.

E. Feather Sorting Test.

When the feathers are received, they are distributed in mixed bunches to the girls in the Raw Feather Department, who are directed to arrange them into piles. In this factory, the different types of feathers are designated as "Amazons," i. e. the broadest and widest feathers; the 18 inch," i. e. those 18 inches long; the "three fourths," i. e.,

those that are as broad as the "Amazons" but shorter; the "Byocks," which are from the tail of the ostrich and are black and white; the "Blacks," i. e. all black, as the name signifies; and "No. 1055," scrawny feathers, thinner and narrower than the others. The girls were told to lay these feathers into the six piles into which they were accustomed to place them. The Firm wished to have a report on the Time Rates and the Accuracy of the girls in this group. Such a test could only be of value when applied to practised factory hands, and was to be used, in this instance to determine which persons should be laid off in the approaching slack period. For weeding out purposes, the test proved to have a distinct value.

Seven ounces of feathers were given as a standard amount: They were laid before the subject en masse. The time for sorting this amount varied from 167 seconds to 329 seconds, with an average time record of 227 seconds; an A. D. of 44.33 seconds; and a S. D. of 11.37 seconds.

There was a range in accuracy from no errors in one instance to fifteen in another. The average error was 4.83; the A. D. was 5.74; the S. D. was 7.29.

Those whose time-score was nearest that of the average, made the best score for accuracy. The very quick and the very slow, had noticeably poor records for accuracy.

A confusion of the Byocks with the Black Feathers, or a failure to discriminate between the other varieties counted as an error and girls were penalized 5 per cent. for each mistake which was deducted from the standard 100 per cent. perfect.

Although the feathers in other factories are occasionally designated differently, the time-records established by this fairly representative group forms a reliable basis for judging the ability of individuals engaged in this type of feather work.

F. Discrimination of Weight and Shape Test.

The girls in the Selecting Department were examined still further by another test designed to determine their quickness of perception and ability to discriminate between small differences in weight.

Eight wooden solids one one-half inches in diameter were so fashioned that there were four pairs, each pair of which had similar bases and sides. Thus there was a pair of seven-sided, eight, nine and ten-sided solids.

These blocks were weighed after the manner of the Binet-Simon weight Test—3 grams, 6, 9, 12, 15, 18, 21, and 24 grams, respectively. There are more blocks in this set than in the Binet, but the same difference between the weight of the blocks has been maintained, that is they were made progressively heavier by three grams.

Two distinct problems were given: (I) Put the blocks in a row according to their weight,—first, the heaviest, then the next heaviest, and so on. (II) Put blocks in pairs, according to the number of their sides. There are two each of like sides.

A time-limit of one minute for each part of the test was allowed. A record of each individual's arrangement of the weights was kept, two correct arrangements out of three were required for credit. Seventy-five per cent. received credit.

The nine and ten-sided figures were most often confused. Fifty per cent. of the girls performed the pairing in 45" without error. The other girls made from two to six errors each and required the entire minute allotted in order to arrive at a decision regarding the blocks with more than six sides. The methods employed by the subjects were enlightening as to the ability of a subject to observe small differences in material, and to arrive at a positive decision regarding those differences.

The girls were graded on the basis of the average rank obtained as a result of the two parts of this test.

Chapter V.

THE CORRELATIONS

The high value of correlations is well recognized and as Prof. E. L. Thorndike has said, "The only sound principle by which one can interpret the tests is determined from the results of the correlations." (57).

The process of obtaining an index of correlation involves the following operations:

First, each subject was ranked One, Two, Three, Four, etc., according to her standing in each individual test, as compared with the rating of her companions in the same test. In the first ten tests the standardized methods of scoring have been employed. The manner of scoring in the special tests devised by the Experimenter is described in the Appendix.

Those subjects who have a like score were ranked in order on the basis of the time required to perform that particular test. That is, if two girls received a rating of 80 per cent. in the Hard Directions Test, for instance, and one of these girls completed the test in 75" while the other girl took 76" to do the same test, the one who was quicker would be ranked higher than her more deliberate co-worker. In case of a "tie" for a given rank, (this rarely occurred), the "ties" were divided in such manner as to keep the total number of ranks equal in the series. If, for example, two subjects ranked 4th in a test, making an identical score in both time and degree of accuracy each was assigned the rank 4.5 in the series thus replacing 4 and 5.

Then, each girl having been given a rank order of merit in each test, a comparative calculation of the inter-relations between each test and every other test was derived by means of the Pearson Formula, in which r, or the coefficient of correlation equals

$$r = 1 - \frac{6 \times \text{Sum of } D^2}{n\,(n^2 - 1)}$$

This coefficient is so derived that when its value is Unity the two variables have perfect concomitance, and when its value is Zero three is an absence of relationship, so far as the conditions of the experimenter are concerned. It is evident that the reliability of this coefficient increases with the num-

ber of cases compared and also with the magnitude of the r obtained.

Seventy-five subjects were tested in this investigation. a sufficiently large number to demonstrate in a fairly reliable manner the value of the tests as indices of the efficiency of the employees.

The Probable Error was determined by means of the Pearson Formula. $.6745 \frac{1-r^2}{\sqrt{n}}$ Unless a correlation be at least twice as large as its probable error it lacks significance, and of course, to be a truly trustworthy index it should be four or five times as large.

The inter-correlations between the tests are indicated in the accompanying table:

But since the results obtained in the various tests administered by the experimenter can have no value until they have been compared with a fixed standard not open to question, it was decided to secure such a standard in the form of an Efficiency Roster which should be compiled from ratings made by the Firm. It will be apparent that a rank order arrangement of the employees made by task-masters who watched

The Tests.	Binet	Trabue	Hard Direc.	Easy Dir. #1	Easy Dir. #2	Av. 3-Dire.	Easy Oppos.	Mixed Rel. #1	Mixed Rel. #2	Subst.	Can. A	Can. 2	Prov.	Motor	Cards	Sort-ing	Color	Feather	Weights	Judg.	Civilian
Binet-Simon	1.	64	69	91	63	73	31	41	43	54	42	40	51	36	40	49					53
Trabue Completion		1.	63	42	60	75	38	61	59	43	25	36	40	47	47	36					62
Hard Directions			1.	75	64	70	32	53	51	40	38	32	57	56	44	44					58
Easy Directions #1				1.	74	76	34	59	56	37	52	42	50	58	40	44					68
Easy Directions #2					1.	63	58	51	35	32	55	31	42	33	36	26					71
Av. 3-Directions						1.	48	54	44	56	47	29	28	60	39	43					70
Easy Opposites							1.	63	42	47	36	42	43	34	58	49					65
Mixed Relations #1								1.	80	33	29	50	23	24	45	42					45
Mixed Relations #2									1.	48	54	47	43	49	44	32					30
Substitution										1.	31	48	50	38	42	40					60
Cancellation -A.											1.	44	28	43	38	35					42
Cancellation Digits												1.	42	31	28	30					69
Proverbs-Eng.-Air.													1.	39	30	47					39
Motor Control														1.	42	54					80
Card-Sorting															1.	56					73
Sorting Test																1.					56
Color Discrim.																	1.	61	26	56	53
Feather Sorting																		1.	37	64	76
Weight-Form																			1.	33	22
Judgment																				1.	65

tion-Firm ...

the performance of these workers for a period of months is a criterion not to be gainsaid. No worker who had been less than six months in the Factory was examined. In order to give more authority to this index, separate appreciations were asked of four different individuals, the Manager, the Foreman and two of the Forewomen.

The procedure was carried on in this manner. The Experimenter arranged four sets of seventy-five cards size 3 x 5 inches. On each card of each set was written the name and the identification number of each of the seventy-five girls who had been submitted to the tests. On the back of each card the Experimenter wrote:—1. General Intelligence. 2. Reliability. 3. Promotion. The four judges were each given a set of these cards and were directed to rank the girls whose names appeared thereon in order of merit; first, on the basis of his personal opinion of the General Intelligence of the girl. The judges were asked to disregard every other characteristic except the intelligence of the individual. The subsequent rankings to be made were not mentioned until each judge had completed his rating. The same directions were given when the other rankings were obtained. After each ranking, the numerical order in which each girl was placed by the judge was noted.

A composite ranking was then derived for each girl by calculating her average position in all four scales in each of three characteristics, Intelligence, Reliability, and rank order for Promotion. Correlations between this Composite or Firm Scale show positive indices raging from plus 28 to 80.

This scale froms the basis for the Teams of Tests to be developed for practical application in the Factory.

Rank Order	Card-Sorting	Substitution	Sorting Test	Cancellation--A	Cancellation of Digits	Motor Control	Criterion Scale
			First Quartile				
1	63	58	29	71	65	66	24
2	46	45	71	1	47	29	47
3	12	29	21	66	13	38	30
4	30	38	50	34	39	45	75
5	49	5	15	38	54	24	65
6	24	25	1	63	20	30	36
7	75	36	46	68	52	47	13
8	13	23	24	9	29	71	45
9	43	49	25	45	75	36	66
10	65	65	66	16	26	65	21
11	68	26	7	3	6	58	32
12	62	66	62	65	1	62	48
13	40	14	65	39	30	43	50
14	36	47	30	42	9	31	25
15	21	24	63	61	48	50	9
16	56	39	13	70	4	32	15
17	47	63	59	36	40	12	59
18	8	71	39	53	66	15	64

Rank Order	Card-Sorting	Substitution	Sorting Test	Cancellation-A	Cancellation of Digits	Motor Control	Criterion Scale
			Second Quartile				
19	45	59	34	29	58	5	5
20	15	30	74	30	32	48	63
21	32	35	16	5	21	25	22
22	9	70	40	10	59	59	58
23	64	32	47	24	5	3	53
24	70	37	68	48	71	22	64
25	50	75	31	54	24	34	16
26	66	72	12	15	50	9	39
27	31	21	75	40	25	21	6
28	22	48	32	59	42	23	37
29	5	31	38	37	43	75	17
30	59	7	72	64	62	5	71
31	37	42	10	50	74	41	72
32	39	52	45	58	17	64	31
33	25	58	28	75	15	68	26
34	71	40	7	43	26	26	7
35	58	53	5	8	68	39	67
36	6	6	37	20	8	72	29
			Third Quartile				
37	60	46	22	23	45	52	46
38	1	67	35	46	63	49	40
39	14	15	3	47	46	16	1
40	17	18	64	55	10	1	54
41	26	16	70	57	12	70	23
42	67	4	58	12	3	14	62
43	41	12	8	17	37	33	35
44	16	3	4	6	50	73	38
45	44	50	42	62	36	46	44
46	3	22	36	4	28	53	60
47	10	1	67	13	41	27	21
48	20	28	52	22	61	11	70
49	23	2	60	33	23	18	74
50	29	8	56	11	19	13	52
51	34	74	41	21	67	63	12
52	38	60	73	52	22	74	49
53	42	10	61	19	38	57	61
54	52	69	53	2	33	42	68
			Fourth Quartile				
55	35	62	11	32	60	8	8
56	27	17	54	25	31	6	10
57	18	19	2	49	18	67	4
58	4	55	43	14	16	69	55
59	28	56	44	60	11	7	11
60	53	9	48	18	70	50	56
61	2	41	69	31	73	60	69
62	72	61	26	51	27	20	43
63	54	64	17	72	7	10	2
64	55	13	55	67	28	35	73
65	19	43	49	73	72	2	42
66	23	27	18	26	69	56	24
67	57	44	19	27	14	7	57
68	7	68	27	56	55	61	27
69	51	57	23	28	35	44	41
70	73	20	20	41	64	28	28
71	48	34	14	44	49	55	19
72	74	73	6	35	44	4	3
73	11	54	33	69	53	37	18
74	20	11	57	74	51	19	33
75	69	51	51	7	57	51	51

Rank Order	Easy Opposites	Mixed Relations No. 1	Mixed Relations No. 2	Trabue Compl.	Proverbs Test	Easy Directions No. 1	Easy Directions No. 2	Hard Directions	Av. 3 Directions	Binet Simon	Rank Order Binet
					First Quartile						
1	59	1	29	39	5	24	71	58	65	65	1.5
2	24	30	70	59	65	36	75	36	75	48	1.5
3	75	24	75	48	39	15	59	75	71	75	5.5
4	65	70	63	5	16	75	15	59	24	45	5.5
5	29	65	38	65	71	59	6	39	36	47	5.5
6	1	39	52	75	66	65	59	70	34	24	5.5
7	15	45	59	58	54	15	65	24	30	71	5.5
8	36	39	7	29	58	63	37	40	5	1	5.5
9	14	16	39	30	74	45	14	36	47	59	10.5
10	71	5	24	17	47	58	48	43	6	29	10.5
11	45	38	66	70	61	47	36	63	15	70	10.5
12	26	16	71	45	24	48	5	5	48	24	10.5
13	58	9	60	21	14	66	21	38	21	25	15.
14	17	29	30	32	34	21	39	52	45	20	15.
15	48	42	1	22	63	6	64	30	58	31	15.
16	70	71	36	61	52	20	7	6	9	21	15.
17	45	40	20	40	43	17	45	48	64	32	15.
18	68	36	16	6	1	37	52	21	7	50	15.
					Second Quartile						
19	30	61	34	36	72	64	46	14	52	6	15.
20	52	59	55	66	38	14	72	29	40	58	15.
21	38	72	61	50	64	46	1	10	66	37	15.
22	20	12	42	52	45	7	49	64	29	46	15.
23	12	60	45	14	17	29	8	15	38	49	15.
24	53	6	37	15	53	54	12	49	1	72	25.
25	32	68	9	71	40	39	55	66	49	62	25.
26	21	21	13	34	36	72	23	74	16	43	25.
27	50	56	72	1	20	40	61	22	46	23	25.
28	63	52	53	67	13	22	58	33	32	9	25.
29	37	74	58	16	8	23	9	20	12	63	25.
30	9	23	6	54	60	50	61	17	23	15	25.
31	72	48	21	60	9	16	66	50	20	56	25.
32	23	28	23	33	48	71	74	7	34	39	25.
33	13	53	17	26	23	12	40	53	42	22	38.5
34	63	41	12	42	26	61	42	34	10	11	38.5
35	31	8	13	46	12	32	4	52	22	15	38.5
33	33	53	17	26	23	12	40	53	42	22	38.5
34	63	41	12	42	26	61	42	34	10	11	38.5
35	31	8	13	46	12	32	4	42	22	54	38.5
36	60	17	28	27	25	31	43	23	74	13	38.5

Rank Order	Easy Opposites	Mixed Relations No. 1	Mixed Relations No. 2	Trabue Comp.	Proverbs Test	Easy Directions No. 1	Easy Directions No. 2	Hard Directions	Av. 3 Directions	Binet Simon	Rank Order Binet
					Third Quartile						
37	35	15	6	28	31	38	50	1	43	42	38.5
38	7	26	50	56	10	8	53	4	61	31	38.5
39	15	22	56	38	68	1	13	37	25	54	38.5
40	34	31	27	8	21	33	19	28	60	52	38.5
41	74	63	43	10	37	28	33	46	72	7	38.5
42	4	7	19	73	56	41	38	54	8	4	38.5
43	25	18	46	43	22	74	32	60	4	55	38.5
44	64	10	73	23	41	11	18	72	26	3	49.
45	42	54	49	12	62	13	28	26	33	40	49.
46	44	3	11	13	28	34	60	62	31	14	49.
47	43	4	15	49	27	67	31	13	56	41	49.
48	11	55	18	57	42	56	50	67	13	27	49.
49	2	19	54	41	35	19	17	61	18	26	49.
50	67	11	74	3	19	26	68	56	67	35	49.
51	41	73	3	68	57	35	41	19	27	12	49.
52	19	32	37	11	3	42	2	8	27	68	49.
53	28	46	41	2	73	27	44	57	55	3	49.
54	57	49	67	4	15	73	10	55	11	28	56.
					Fouth Quartile						
55	8	13	14	53	34	68	74	44	41	60	56.
56	61	64	25	20	32	18	56	73	68	10	59.
57	3	67	4	74	32	55	25	41	73	68	56.
58	18	2	48	35	4	44	51	11	2	56	56.
59	70	71	20	61	43	22	7	6	29	51	59.
60	56	69	27	25	54	62	35	69	42	19	59.
61	73	33	4	44	18	3	57	27	3	57	59.
62	69	57	2	69	51	69	3	35	51	18	59.
63	38	72	61	50	64	46	26	14	37	49	59.
64	10	14	35	55	46	4	67	31	2	28	68.
65	40	54	41	12	62	13	28	62	33	43	68.
66	8	13	14	50	44	68	74	41	19	60	68.
67	35	10	6	28	31	38	53	4	61	61	68.
68	71	55	44	62	60	57	11	2	57	44	68.
69	73	30	4	44	18	3	69	30	51	2	69.
70	19	32	31	11	73	42	2	8	27	33	68.
71	5	20	69	25	54	62	62	69	46	69	68.
72	6	25	28	62	69	57	35	2	57	44	72.
73	16	27	33	44	18	3	57	27	3	57	73.
74	22	34	57	51	55	69	3	28	51	18	74.
75	27	51	20	69	44	33	20	51	69	51	75.

When the members of the Firm were asked to arrange all the employees in rank order on the basis of their suitability for Promotion, they stated that some persons would not be promoted under any consideration. All four judges agreed that twenty-four out of the seventy-five employees would remain statu quo. The twenty-four cards representing

these persons were removed from the pack. The Manager then consented to arrange this collection of non-promotable persons in rank order on the ground of their general efficiency. According to this ranking, the twenty-four employees designated were given grades after the fifty-one had been placed in order for promotion. The Judges arranged them also in rank order on the basis of the other two characteristics and the average position thus derived placed them in gradation with their confreres.

It is apparent that the four judges may disagree as to the exact rank of the individual girl; their ideas as to what is meant by the terms General Intelligence and Reliability may be unlike, but the average position achieved by each girl in these three characteristics cannot be biassed by individual misapprehensions. Ranked order scales in estimating the efficiency of a large group of persons are admittedly fairer when a combined rating of several competent judges is made.

By an inspection of the Table of Correlations, certain facts of interest may be noted.

The easiest tests bear a higher relation to the Criterion than do the more difficult tests such as the Proverbs and the two Mixed Relations Tests.

The results in the Mixed Relations Tests do not confirm the findings of Wyatt (51) who states that they form the highest correlation with Intelligence of any of the tests save the Trabue Completion.

The Coefficients of Correlation between the Special Tests and the Criterion demonstrate that the Card Sorting, the Sorting Test No. 2 and the Motor Control Test are of value, and inasmuch as they represent motor activities similar to tasks performed in the Factory they should be regarded with special significance. The Card Sorting Test gives an index of plus .73., P. E. .0280. The Sorting Test No. 2 a positive index of .60., P. E. .0498. The Motor Control Test gives the highest coefficient of any of the tests that were presented to the group, i. e. plus .80., P. E. .02484.

Both of the Mixed Relationship Tests show fairly high indices of correlations between them and the Directions Tests, and between them and the Linguistic Tests. In studying the Table the low inter-correlations between these two tests and the purely mechanical tests are not surprising. In the records of individual girls it is apparent that with a few exceptions some of them do consistently well in all tests that demand on or the other type of ability, scholastic or motor.

The Cancellation Tests correlated with each of the other tests give low but positive indices. As was noted under the discussion of the Binet Test, this type of test does not bear a high correlation with tests that demand General Intelligence.

Dr. Weidensall found a positive index of .52., P. E. .055. between the A-Test and General Intelligence (31); while our coefficient is but .36., P. E. .0681. In the same group of Bedford women tested, she found a correspondence in this test between the rate of its performance and the difference in intelligence as indicated by school grade. This was to be noted in the Factory group also. By a comparison made between the subjects arranged in order on the basis of their time-records in the A-Test and the grade in school attained at the age of fourteen years, one finds the retarded pupils required 7 per cent. longer than did the median of the group to perform the test.

Terman and Bagley obtained a negative correlation and Simpson (59) a slightly positive one between intelligence and skill in the A-Test.

In "School and Society" (Vol. V., P. 24), there is a report of an experiment in which the Cancellation Tests correlate negatively with the Composite Tests which are adjudged good measures of mental acumen. To quote the statement of Dr. McCall, "This proves that a negative correlation may exist between apparently desirable traits."

In this study, however, the Cancellation Tests bear positive rather than negative relations to the other tests and would seem to uphold Prof. Thorndike's (60) theory that there is not an inverse ratio between desirable traits.

Correlated with each other, the Cancellation A-Test and the Digits Test indicate a relationship of plus .44., P. E. .0629. This is not, comparatively speaking, a high correlation for between tests that demand the same type of reaction there is usually a closer index of relationship. For instance, between the two Mixed Relationship Tests there is a positive coefficient of .80., P. E. .0280., and between the three Directions tests there are fairly high indices of correlation.

The reason for this case of low inter-correlation between two tests of a similar type may have been a lack of interest in continuing at work of the same kind, evidenced by several persons when performing the second cancellation test.

From an inspection of this table, there are no indications of such a hierarchy of coefficients as is demonstrated by Burt's "Hierarchy of the Specific Intelligences".

Ranked in the order of the degree in which they correlate with the Criterion, the tests are Motor Control, the Card Sorting test; Easy Directions No. 2, Averaged 3-Directions, Easy Directions No. 1; Easy Opposites; Trabue Completion, (1), the Substitution Test, Sorting Test, Hard Directions, Binet Proverbs, Digits, Average of 2-Cancellation Tests, Mixed Relations No. 2., A-Test.

Although there is no very special conclusion to be drawn in regard to the possession of "general ability", the more intelligent of the subjects tend to do well in whatsoever test they are given to execute.

The Special Tests.

Twenty girls chosen at random from the Selecting Department were submitted to the following tests especially devised by the Experimenter to measure their capacity to perform the tasks demanded in this department: (1). Weight and Form; (2). Color Discrimination; (3). Feather Sorting; (4). Judgment. These tests are described on Pages 50-60.

Before proceeding with the tests, the Manager, the Foreman and the Forewoman in charge of the Selecting Department made a rank order arrangement of this particular group of employees on the basis of the "Efficiency displayed by each in their daily work". The average of these three ratings formed a new scale, the Criterion, with which the girls' standing in the four tests have been correlated.

Rank Order of Selecting Department Girls in Tests indicated. Criterion—Three-fold estimate of Firm.

Rank	Color	Judgment	Weights	Feathers	Criterion
1.	21	28	63	24	63
2.	53	24	33	74	24
3.	28	46	28	49	28
4.	24	63	21	63	68
5.	74	68	12	21	64
6.	63	53	49	68	21
7.	50	21	64	28	50
8.	12	64	8	46	46
9.	49	12	74	8	74
10.	64	8	41	53	1
11.	8	49	71	1	12
12.	46	74	50	41	49
13.	71	33	68	12	16
14.	16	71	53	50	53
15.	69	51	24	64	69
16.	41	1	69	33	41
17.	68	16	1	69	8
18.	1	69	46	16	51
19.	33	50	51	71	71
20.	51	41	16	51	33

a The coefficients were derived by means of the Formula

$$r = 1 - \frac{6 \times \text{sum of } D^2}{n\,(n^2 - 1)}$$ In this study $n = 20$.

The result of this correlation is shown in this table:

	Wt. Form	Color Dis.	Feathers	Judgment	Criterion
Wt. Form	1.00	.26	.37	.33	.22
Color Dis.		1.00	.61	.56	.58
Feathers			1.00	.64	.76
Judgment				1.00	.65
Criterion					

The Feather Sorting Test which involves the use of the material the girls are familiar with gives the highest coefficient of correlation, .76.

The Judgment Test which the Firm found of so much value that it was adopted immediately by the Employment Foreman, rank second. This test measures the capacity for making quick and accurate decisions, and is also similar in its demands to the work of the department.

Time vs. Accuracy.

One of the differences between a superior and inferior worker lies in the speed with which each performs his task.

In studying the individual variations of the members of this group two fairly distinct types of time-scores may be noted the time-record of the slow, methodical worker and that of the nervous, quick, distractable worker. The deliberate and painstaking girl stood higher in the estimation of her employers than did the quicker and more alert girl who was inclined to be inaccurate and careless. The latter made good time records but in excellence of performance, her score was low. Many of these girls over-emphasized the desirability for haste and rushed through the tests, assuming that the faster they worked the better it would be for them. This was done regardless of the admonition to "work carefully". Such reactions were noticeable in the first written tests performed by Subjects No. 25, No. 56, 60, and No. 70. in the Easy Opposites and in the Mixed Relations Tests. Speed and Accuracy with them revealed a negative correlation.

The reactions of the employees numbered 65, 75, 24, 71, 15, and 63 for example, showed a fair and consistently high rating in the tests so far as accuracy or performance went but their time-scores were variable and not below average length.

The Coefficient of Correlation between Time and score for this group is usually highest in those tests in which a demand is made primarily upon motor activity. Thus, in the

following tests there is a positive correlation between time and score of:

Test	Coefficient	P. E.
Substitution	.44	.028.
Digits	.57	.0361.
Sorting Test	.62	.031.
Motor Control	.65	.019.

In the Linguistic Tests, the indices between Time and Score show positive but low correlation, thus:

Trabue Completion—.23., P.E. .081.; Proverbs Test, .20., P.E....015.

The Directions Tests give positive indices of .50., 51. and .53., respectively.

Between Time and Score in the Mixed Relations Tests there was a positive coefficient of .27., P.E. .037., and .31., P.E. .036.

While it cannot be considered conclusive evidence that the correlation between Speed and Accuracy is always higher in tests involving motor activity when such a group as this is examined, the probability is that this would be the case.

The question arises as to the value that should be placed upon Speed alone as a legitimate index of mental efficiency. In the Factory, Speed per se has, in general, a negligible appraisal. Unless Speed is accompanied by Accuracy, the Firm announced it an undesirable trait in an employee.

It is no doubt true that when some individuals, working under constant conditions, increase their speed, they tend to increase the number of errors, but other individuals under like conditions may show a decrease in their errors. Prof. Whipple (7) quotes results that show that the faster subjects are also the most accurate. In six of his groups, Brown (58) found positive correlations between Speed and Accuracy of Adding that ranged from .13 to 43., with P. E.s ranging from .07 to .12. With small groups of college students, Prof. Whipple states that he has obtained positive correlations of .19 in the case of adding and as high as .86 for mental multiplication. (7).

The Factory Girls cannot be judged by the same standards as the College Students in the reaction time required to perform mental tests.

All save those subjects whose standing in the tests placed them in the upper third of the group react slowly to all stimuli. Their cerebrations are deliberate even when the simplest matters are under consideration. In some cases, the slow reactions are due to the fact that the subjects are unaccustomed to writing, and the co-ordinations necessary for

this form of activity show disuse. Though the median age of the subjects was seventeen, there are several of the older women have been out of school a great many years. For them the task of writing their responses involved an actually laborious application of unhabitual motivity.

The weight, then, that is to be placed upon an individual girl's time-record in her performance of tests in this Factory must depend upon her accuracy of achievement rather than upon the time element alone.

Chapter VI.

1. MULTIPLE CORRELATIONS.

The correlations obtained between the results of the employees' standing in each of the several tests and the ranking of these employees made by the Firm vary according to the particular test applied. However, in the sum-total of reactions they indicate that the verdict secured through the tests approximates the valuation established by the Firm especially as it affects the upper and lower quartiles of the Scales involved.

Meumann (59) states that it is an error frequently made by the exponents of what he calls "correlation psychology" to conclude that because a high correlation exists between any two elements it is necessarily a proof of the good qualities of both. "Both", he states, "may be one-sided". A combination, therefore, of several tests should be made in order to obtain their composite valuation. Thus any criticism that too great emphasis has been placed upon the result of the correlations between the Criterion and single tests may be forestalled.

Stern (60) and Burt (61) found that the higher correlation of the combined results of several tests show that one test compensates for another, each calling forth particular mental functions, so that only through such combination can a total picture of intelligence be created. The combined coefficients obtained from several tests have been calculated, therefore, in this group.

Indeed, one of the objects of this investigation is to demonstrate the combinations of tests that may be given practical application at the Factory. For it is an obvious fact that the index derived from the presentation of a group of tests will be a more reliable indicator than will the single isolated test that is naturally more or less narrow in scope.

The selection of tests to form a Team should be made on the basis of the magnitude of the coefficients of correlation.

"A Multiple Correlation Coefficient is that correlation which expresses the total efficiency of the scale when the tests choosen are those that bear the best or highest correlation with the Criterion." By this method there may be constructed a combination of tests which will predict "general ability" better than can be done by a single test.

Tests for a Team should bear high correlation to the Criterion and low indices of correlation with each other.

High inter-correlations indicate the existence of a too-close resemblance between the tests to make their combination and subsequent use as a Team worth while.

Recently, Dr. Herbert Toops of the Institute for Educational Research has devised a new method for the more rapid calculation of Test-Combinations than those formerly in use. He has transmuted algebraically the old method of procedure and by means of a printed form made it possible to reduce the formula, $\frac{(ric)(riu) \quad 2(ruc)(ric)(riu)}{1-(ruc)^2}$ to a series of more or less mechanical operations. This method has not yet been published, and I am indebted to him for the verbal demonstration that has already facilitated this portion of the present study.

In general terms, the rule is to select from the Table of Correlations that test which bears the highest relationship to the Criterion. Ordinarily, the subsequent choice of a test should be one which bears a high correlation to the Criterion but a low index to the first test chosen.

We assume a weighting of 1.00 for the first test as its Coefficient more closely approximates unity than do any of the other tests. The object in selecting a Team of Tests is to determine that series of tests which will approach most nearly to the ideal relationship of unity.

Following the prescribed method of procedure, test after test may be added to the original pair and an increasingly higher coefficient be obtained without changing the weights of the tests already in the combination.

Occasionally, the gain in the increment added is not sufficiently large to make the particular test combined of value. When this occurs, that especial test is to be eliminated so far as that test-combination is concerned and another test may be chosen in its place. Herein lies one of the apparent advantages offered by this method.

As an illustration, in this study, the Easy Directions Test No. 2 with its positive coefficient of .71 was selected as the Key Test; to it was added the Substitution Test, with an index of .60 with the Criterion and its inter-correlation of .32. This combination gives an index of .8368. Next, the Digits Test was added with its Criterion Coefficient of .49., and its inter-relating coefficients of .31 and .42 which again raises the multiple coefficient to .8432. By adding the Sorting Test (Criterion Coef. equals .56., inter-correlations equal .30., 36.) to this combination, the coefficient becomes .8436., a gain of only .0013. Hence, it is advisable to drop the Sorting Test

in this particular combination, and in its place, add the Cancellation of A-Test which has a Criterion coefficient of .39. The Multiple Coefficient then becomes .8747. As these tests require as a maximum about ten minutes for their execution, the Team seems an extremely practical one.

The next combination selected consists of the Card Sorting Test with its coefficient of .73 with the Criterion and the Average of the Three Directions Tests which correlates with the Criterion with a coefficient of .70. The inter-correlation between the two is .55. Combined, the two tests give a coefficient of .8128. Now if we add to this combination the Trabue with its coefficient of .62 when correlated with the Criterion, and of .75 correlatd with the Three Directions, and of .47 correlated with the Sorting Test, we get an increase in our coefficient of .0049; to this combination, add the Digits Test (Criterion Correlation equals .49; Inter-correlations of .35., 42., .43.,) which raises the coefficient to .8217; to this we now choose to add the Substitution Test which has a coefficient of .60 when correlated with the Criterion, and of .56., .48., .43 and .40 when correlated with the rest of tests in the combination. The coefficient resulting from this operation rises more decisively and gives us a coefficient of .8507

Judging from the time-records made by this group of girls, this entire set would require approximately 22' 48.5". Some of the other test-combinations have proved of greater value because less time is required for their execution, and they offer a larger variety of mental exercises.

For instance, the Motor Control Test which combined with the Sorting Test (Criterion Coefficients of .80 and 73; inter-correlation of.54) gives an index of .8667. Adding to it the Easy Directions No. 2, (Criterion Index equals .71., inter-correlations equal .36: .33), and the Multiple Coefficient becomes .8773., an increase of .0106. If the Easy Opposites be added, (Criterion Coefficient equals .65; inter-correlations equal .49., 34., .68), the Coefficient becomes .9292. This is a good practical combination inasmuch as it may be administered within a limit of 500", and combines variety and brevity.

Another combination consists of the Easy Directions No. 2 and the Card Sorting Tests (Criterion Coefficients of .71 and .56 and an inter-correlation of .34) which gives a coefficient of .7867. To this, add the Digits Test (Criterion Coefficient equals .49; inter-correlations of .33 and .31) and the Multiple becomes .8118. Then add the Substitution Test (Criterion Coefficient equals .60; inter-correlations equal .32; .43; .48) and our Multiple Index becomes .8409. These tests require about 675." and have already proved of value in practical application at the Factory.

The next group is comprised of the Card Sorting and Three Directions Tests Combined, (Criterion Coefficients

equal .70 and .73 and inter-correlations equal .58); together these two tests give a coefficient of .7812. When the Easy Opposites Test is added with its Criterion Coefficient of .65., and inter-correlations of .58 and .45, the coefficient becomes .8327. Digits added to the series increases the coefficient to .8544. This combination requires about 807" for its execution.

Another combination is comprised of the Motor Control and the Easy Directions No. 2, which give a Multiple of .9071; Sorting adds .0494; Substitution .0160; and Digits .0027: making a total coefficient of .9681., the best combination that has been derived. It is composed of tests which can be given in approximately 13 minutes. Each one of these tests has been given a practical trial by the Employment Managers and they are continuing to prove their worth as a rapid means of predicting the efficiency of an employee.

The Special Tests devised for use in this particular industry were next combined by taking first the Feather Sorting the Judgment Tests with their Criterion Coefficients of .76 and .65 and their inter-correlation coefficients of .64, whereby the Multiple Coefficient becomes .7944. When the Color Test is added to this group the coefficient becomes .8002. The Weight Test make no contribution of sufficient size to warrant the time required for its presentation.

It is to be noted that in determining these Teams, only tests bearing the highest indices of correlation with the Criterion have been considered. Thus, the two Mixed Relations, the Proverbs and the Weight Tests have been dropped. So far as this type of work is concerned they do not predict failure or success, their contribution is practically nil. The Binet requires too much time to be practical; only the single tests that have been designated elsewhere are to be retained.

To ascertain the weighting of the tests retained, the tables and procedure arranged by Dr. Toops were applied.

The first step is to choose that test which bears the highest correlation to the Criterion. Each test in turn is combined with the first test to ascertain what the special contribution of each to the combination already determined may be.

We take, then, the Motor Test as the "combination" when the Binet is combined with it, the Binet has a weighting of .6736; the Motor and the other tests have a weighting as follows:

Motor and Trabue	give	a	weighting	of	.4797.
Hard D.	"	"	"	"	.2778.
Easy D. No. 2	"	"	"	"	.7398.
3-D. Av.	"	"	"	"	.4146.
Easy Opp.	"	"	"	"	.2921.

Mixed Rels. No. 1	"	"	"	"	.2751.
Mixed Rels. No. 2	"	"	"	"	.0003.
Substitution	"	"	"	"	.5175.
A-Test Cancel.	"	"	"	"	.1347.
Digits	"	"	"	"	.3735.
Proverbs	"	"	"	"	.1439.
Cards	"	"	"	"	.7137.
Sorting	"	"	"	"	.2572.

By inspection the Easy Directions No. 2 and the Cards Test have the highest weighting. The first two combined with the Motor Test make a contribution of .8968,; while Cards and Motor have a correlation of .8666. Substitution with Motor as a combination have a correlation of .8616; the contribution of any of the other test is of less value at this point.

Taking Motor and the Easy Directions No. 2 as the combination, on the basis of the weightings already determined, we add the Trabue, which at this point receives a weighting of .1880; the Hard Directions has a weighting in this combination of 1419; Easy Directions No. 1 of .0169; Easy Opposites of .1201; Mixed Relations No. 1 has a weighting .0321; Mixed Relations No. 2, a weighting of 1363; Subsitution has a weighting of .5162; the A-Test of weighting of .1828; and Digits has a weighting of .3348; Card-Sorting, a weighting of .9390; Sorting Test a weighting of .1767; The highest of these weightings, the Card-Sorting Test, is, selected for the next combination.

With this new combination, we then derive the weighting of each of the other tests, in turn. Thus, when Motor, Easy Directions No. 2 and the Card tests are combined, the Trabue contribues .0823; Substitution .4734; Digits .3382. The other weightings are of less value. When Substitution and Digits are added to the former Combination, the aggregate correlation is .9591. Each test makes some contribution to the whole weighting, but in this case, additional tests increase the sum-total contribution in only a slight measure.

The Special Test should now be considered. Feathers and the Color Tests combined give a weighting of .2866; Weights test has a negative value of .0903., and is, therefore, to be dropped. When Judgment is combined with Feathers and Color, the Feather test remaining as the constant, the weighting becomes .4756. The correlation then becomes .7893. With Color added, the correlation becomes .8002.

Surveying all the weightings in this group of test, the contributions made by the Easy Directions No. 2., the Substitution, Cards and Digits tests, if the Motor Test is chosen as a key test, have the greatest value.

Of the Special Tests, the Feather Sorting and the Judgment Tests are to be selected in preference to the Color and Weights tests.

It is possible, therefore, by this method to determine the actual contribution made by each test, and let the weighting thus derived be a guide as to the retention or rejection of the tests under consideration.

Before practical use of these methods can be made, however, it is necessary to predict the scores within the limits of which the test-records of the subject must lie.

The first step in this procedure is to select those teams of tests which bear the highest coefficient of correlation to the Criterion and contain the individual tests which prove of the greatest value in their practical application.

It is essential to choose more than one team because of the varied demands of the Factory work. Hence, teams No. 2., No. 3., No. 6 and No. 7 are chosen.

The Teams of Tests which are to be applied in this industry.*

Team II.

Trabue Completion
Sorting
Three-Directions
Digits
Substitution

Team III.

Motor Control
Sorting
Easy Directions No. 2
Easy Opposites

Team VI.

Motor Control
Easy Directions No. 2
Substitution
Sorting
Digits

Team VII. Special Tests

Feather Sorting
Judgment
Color Discrimination

It is of practical value to have a definite method for scoring each Team of Tests. For this purpose it is necessary to obtain the combined predicted score for each Team. This is determined by dividing the Beta Weighting (obtained by the method demonstrated above, of each test by the Standard Deviation derived in that test, and multiplying each individual's test-score by the quotient thus obtained. This resulting quotient is usually an unwieldy number, a decimal that is difficult to manipulate. It is, therefore, advisable to reduce all the decimals in the Team to a workable form. This can be done by either dividing or multiplying each quotient by the same number. The algebraic theorem which states that multiplying or dividing all the members of an expression by the same number does not alter the existing relationship

*The directions for giving the special tests are found in the Appendix

between the members, justifies this proceeding. Having determined by which process to reduce the quotients to numbers that will make convenient multipliers the next step in the procedure is to multiply each girl's test-score by the multiplier derived for each individual test. The sums of all the products is the total score obtained by that individual in the tests of that particular Team.

The range of test-scores varies. The significant value of this operation lies, therefore, in the fact that, all the test-scores are reduced to one level and a subject is said to be rated in a certain decile in the combination of tests, rather than in one decile in one test and in another decile in another test etc.

Now, in order to determine what may be the level of achievment demanded in Team No. 2, for instance, it is to be noted that since the possible scores range from approximately 4224 to 1418 as a combination record, these are to be regarded as test-limits. The accomanying table may be used as a sample illustration of the method.

The girls whose individual records range from approximately 3000 to the upper limit in Team No. 2 do excellent work; those whose records range from 2000 to 3000 are fairly good, while those below the 2000 limit are doing poor work. Therefore, we are justified in demanding a combination score in Team Number Two of approximately 55 per cent. of the highest score made.

The same method of procedure is to be followed in each of the other Teams.

Thus, the highest combination score for Team 3 is approximately 5000 and its lowest limit is about 1300. We find no subjects on this scale who are doing good work, whose records are below about 2500, which is approximately 60 per cent. of the average record of the highest Group in this test. In Team 3 the multiples are 19 for the Motor Test; 12 for the Sorting Test; 8 for Easy Directions No. 2 and 7 for the Easy Opposites.

In Team 6 the multiple for the Motor Test is 19; for the Easy Directions Test is 14; for the Card-Sorting Test is 9; for the Substitution Test is 3; and for the Digits Test is 1.

In Team 6 there is a range of possible scores of from approximately 1500 to 4000. In this case we shall call 60 per cent. of the average score of those in the highest quartile of the test a good passing mark. Thus, those whose record scores are below an approximate rating of 2500 will not be valuable assets in the Selecting Department.

The Special Tests, of which only 3 are retained, the Feather Sorting, Judgment and Color Tests, have these multiples respectively, 4, 6 and 3.

The range in this test is from approximately 1200 to 700 as limiting scores.

In this instance, those obtaining a combination score of approximately 1100 and over are to be rated as excellent, those whose scores range from 900 to 1100 are fair, and any person grading below the level of approximately 900 are to be ranked as poor. This group is to be given a 75 per cent. pass mark, or an obligatory score of approximately 900 points.

A record blank can readily be arranged upon this basis for each team of tests. . .

FORM TO BE USED FOR RATING SUBJECTS

Name	Fitness	Credits			Total
	Test 1	Test 2	Test 3	Test 4	Scores
Maximum Scores					
	100%	97%	90%	100%	

PERCENT RATING OF 5 SUBJECTS IN TESTS OF TEAM VII.

No.	x 4 Feather Test	x 6 Judgment	x 3 Color	Total
24	100	95	86	1228
49	95	80	80	1100
12	75	82	83	1041
50	70	72	82	958
51	25	74	60	724

Chapter VII.

INTERPRETATION OF THE DATA.

No form of measurement, whether it be physical or mental can be absolutely exact, but the more tests we give, the closer is our approach to a perfect index of the subject's mental status. The large number of tests applied in this investigation was made because of a desire to obtain as complete a mental picture of the group as was possible.

The art of evaluating tests scientifically consists in making checks and counter-checks that it may be definitely determined to what extent the test-results may be depended upon. This has been done in our study by the methods of correlation. The results demonstrate (1) which tests are of greatest value from the view-point of the employer; (2) how tests compare in difficulty of performance; (3) what combinations of tests reveal the best standards by which to judge an employee's capacity for work in the factory; (4) the relative standing of individual girls; (5) the relative standing of the groups engaged in the different departments of the factory; (6) the indices of test-values derived from comparing the rating of the girls in the tests with their rating by the Firm as revealed in their daily work.

The aim has been to simply register all the positive indications of intelligence. By means of the test-results, different levels of proficiency requisite for a person who is to be engaged in any one of the different departments of this particular type of factory have been established. The "levels" vary with the demands of the several departments.

In the second chapter of this investigation, the work required in each of the departments of the factory is analyzed. This study would be incomplete if the analysis of the requirements of the job and of the proficiency of the girls as established by the tests were not brought together. Subsequent chapters have been devoted to the analysis of test-records and a comparison of those records with the rankings of the firm.

The results of the application of methods to determine the Partial and Multiple Correlations established between the tests and the Criterion are in themselves sufficiently significantto require no further comment regarding their value. The "Teams of Tests" chosen by these methods of selection

and elimination demonstrate the best means by which to secure a swift and accurate picture of the applicant's native capacity, in so far as that capacity will prove of use in this type of work.

In the Fancy Department, it has been demonstrated (Page 42) that the ability to reproduce the two designs included in the Binet Scale is a highly desirable accomplishment for its workers. The Team of Tests 2 should be given; 60 per cent. must be obtained. If a subject falls below this standard the chances are that she will be more satisfactory in one of the other departments where a low grade of mentality is acceptable.

In the Selecting Department, good judgment and the ability to continue at a monotonous task are best revealed by TeamVI and the Binet Weight Test. Grade C in the memory tests on the Binet Scale is also a requisite for the best grade of worker. Team VII should be given to determine which persons possess special ability for this type of work. The Selectors of Raw Feathers do not need to reach as high a grade of intelligence as do the Selectors of the Finished Feathers and Team III should be given to them. The 3rd or 4th Grade on the Motor Control Graded Scale is sufficiently good. Grade C in the Binet Memory Tests is desirable. There should be no more than seven piles in the Sorting Test, and the time for the performance should not exceed 360".

In the Boa Department, the Motor Control Test is again the principal criterion by which to judge whether a person can engage to advantage in a monotonous task. Team III or Team VI should be given. Grade C in the Auditory Memory Tests is sufficient.

In the Dyeing Department, Team II is to be given.

The Benders and Twisters should be given Team 2 graded 3rd or 4th in the Motor Test.

The Errand Girls should have B-Grading in the Auditory Memory Tests from the Binet-Scale and be given Team II.

In this schedule it will be noted that the lowest standard in test-rating is acceptable in the Sorting of Raw Feathers, and in the Department of Benders and Twisters. It was found that many girls in these two departments were below average intelligence. None of them had a mental age on the Binet-Scale of more than 10.8., and the majority ranged from 9.4 to 10., while one girl (151) measured but 8.6. The point of importance in this connection is that great many persons who rank below the norm in psychological tests could be usefully employed if a careful study were made not only of the mental equipment of the subjects but also of the degree of mental development required for the work in factories.

Until we know by what mental scale the majority of those engaged in Factory Work should be judged, we cannot

determine any definite criterion for the mental measurements of the steady, unskilled workers of the world. In this study, if we consider the Binet Tests apart from the other standards of measurement, a condition is revealed which permits the conclusion that a great many individuals often regarded as a dead loss to the industrial world can be utilized to good purpose in a number of industries where manual dexterity is more of a requirement than judgment or the ability to reason.

The Experimenter believes that this study has laid the foundation for the utilization of Teams of Tests calculated to measure this grade of labor material.

Furthermore, other questions whose solution will greatly increase the efficiency of the factory have been solved. The tests have determined whether an applicant will do better work as a selector, a bender, an employee in the Fancy Department, in the Dyeing Department or as an Errand Girl.

When promotions are to be made, the Standards of achievement that are to be considered a legitimate gauge by which to designate the worthiest candidates can be ascertained from an inspection of the analysis of the requirements set forth in the Teams of Tests. Teams II, VI, and VII should be given to those eligible for Promotion.

The fact that the relationship existent between the tests and the Criterion are of a positive nature, and that, in some cases they are of considerable magnitude, are convincing proof of the practical applicability of the methods delineated.

Chapter VIII.

CONFIRMATION OF THE TESTS
Three Years Later.

Calculations in which the human element enters to the degree in which it is present in such tests as those described in this paper are always open to question by the non-scientific who are prone to look upon this sort of analysis as theory pure and simple.

In order to refute any possible charge that the actual progression and evolution of the personnel of the Factory had been in contradiction to the results of the tests and the conclusions drawn therefrom, the Examiner visited the factory after allowing an interval of three years to elapse and carefully ascertained the status of progression or regression of every subject examined three years, earlier who still remained in the employ of the Firm. As much information as possible was gathered regarding those who had left the firm.

It is a source of satisfaction to record that the findings of the tests are vindicated by a fairly close parallel between what the tests predicted and what actually developed in the operation of the factory. The girls scoring the highest mental rating were found to have won promotion to positions of Forewomen, while at the other end of the scale those rating lowest had been dismissed and were no longer employed, or else had remained in statu quo.

Twenty-eight of the original force were found to be still in the employ of the Firm. Of the forty-seven persons who had left, however, thirteen had married, so that they are not to be included in the group which the tests designated as lacking in stability.

It is interesting to note that the mental status revealed by the tests of the subjects leaving to get married is contradictory and inconclusive and permits no such deduction, for instance, as that the more mentally alert have a better matrimonial chance than those ranking lower in the scale. The records show that the subjects achieving matrimony were nearly equally divided between the four quartiles of the scale.

There were, then, thirty-four girls who had left the Factory for other reasons than matrimony.

Of this thirty-four had proved undesirable or undesiring about half had been designated by the tests as unsuited for monotonous work. The Motor Control Test was the principal

index in this diagnosis, although the careless execution of the two Cancellation tests was also noted in this connection.

The distinction between undesirable and undesiring has been made because in the class of unsteadfast persons are some who are in no way mentally lethargic, but are, on the contrary, alert and adventurous, and leave jobs, not through inefficiency but through a craving for larger opportunities. Persons such as No. 48., No. 15., and No. 5. were, it was ascertained, employed as saleswoman, telephone operator and cashier in other concerns. All three of these occupations put a girl on a somewhat higher social basis in the estimation of her comrades, and are also jobs offering an increase in salary.

Subjects No. 2., No. 7., No. 11., No. 27. and No. 42 had left the Factory through other causes than matrimony, or an advance in position. Their performances in the Motor and Cancellation and Sorting Tests shewed them to be incapacitated for monotonous labor and this fact was incorporated in the reports made at the time they were tested.

An Employer taking into consideration the results of the Motor, Cancellation and Sorting Tests would never have given work of a monotonous character or tasks requiring delicate manipulative ability to girls ranking as these girls ranked under tests designated to bring out manual dexterity and the capacity to persevere at a repetitive type of work.

A comparison between the records of the 28 girls who had remained at the Factory and the records of those who had left shows that the first group measured, on an average, 11.11 on the Binet-Simon Scale, while the second group scored an average mental age of 10.85, according to the results of the Binet test. In the Association, Linguistic, Directions and Special Tests, the first group obtained an average score of 73 per cent. with a mean deviation of 6.75., while the second group made an average score of 69.56 per cent., with a mean deviation of 8.9. The differences, though slight, are suggestive and might be of greater significance in a larger group.

The original testing established also what three years of subsequent experience confirmed, namely that subjects No. 65., No. 3., No. 63., No. 24., No. 45 were capable of performing the tasks assigned to them. These girls were found to have progressed in value to the Firm, and in productiveness to themselves.

The most promising subjects, as revealed by high test-scores, of those originally examined, include No. 65., No. 29., No. 24., No. 71., No. 5., No. 75. These subjects never fell below the median in the test-scores in both the Motor and Intelligence tests. They were found to have been advanced to positions of responsibility demanding initiative and alertness of mind. Subject No. 65 is often called upon now to

converse with the Firm's customers and demonstrate the finished product placed on sale.

Subjects No. 24 and 30 who obtained ratings in the first and second quartiles in the tests were occupying positions as Forewomen. Numbers 46., 23., 20., 67., 12., and 42. were continuing the same work in which they had been engaged at the time of their examination, but their wages had been advanced owing to their efficiency and increased productiveness.

Twelve of the girls remaining in the employ of the Firm upon this re-investigation were found to have been graded generally in the 3rd or 4th Quartiles in their Test-Records.

To those who have made a study of psychological testing, it will not be surprising that such corroboration of the test-findings was obtained both at the top and at the bottom of the Examiner's scale. As is well-known, the tests offer a reliable means of ascertaining the very quick and the very slow.

Not only, however, are the tests of high value in assisting the Employer to arrive at a quick verdict as regards those ranking at the two extremes of the scale but what is even more desirable they provide for him a means of arriving at a fairly accurate estimate of the industrial value of those in the zone between the very good and the very poor.

APPENDIX.

SPECIAL TESTS FOR THIS INDUSTRY.

VI. 1. Judgment Test.

Twenty strips of card-board were cut to approximate the dimensions of the feathers used in the factory, thus:

Feather Dummies.

LENGTH		WIDTH	LENGTH		WIDTH
20.5 in.	x	9.75 in.	12. in.	x	7.25
17.	x	8.	13.5	x	8.
8.5	x	7.5	15.5	x	8.25
18.	x	7.	14.5	x	17.25
8.5	x	6.125	11.	x	7.
12.5	x	6.5	11.5	x	7.
16.5	x	10.5	14.	x	8.75
18.	x	11.	16.5	x	8.25
18.25	x	9.	10.	x	9.25
15.5	x	8.	22.5	x	11.

Instructions: Each strip of card-board should be cut accurately as per above scale from fairly heavy card-board. An extra strip should be shown to the subject. The subject is told that a number of pieces of card board are to be held before her and that each one has been cut to correspond to the length and width of the feathers which she is in the habit of sorting. The experimenter says further: You are to observe the strip held before you and you are to write your estimate of the length and width of each feather-dummy shown you. There are numbers on the sheet of paper given you that correspond with the numbers on the pieces of card-board.

Each individual was given a sheet of paper at the top of which she was told to write her name and the number that had been given her. Below, under the word "Length", she was to write her approximate estimate of the length of the card, and under the word "Width", the estimated width. A sheet of paper with these headings was given to each girl.

Name.....

Identification Number.....

Length.....Width

1. by

2. by

3.....etc. up to and including 20.

The pieces of card-board were each held at a distance of 18 inches from the subject and exposed for 10". A stop-watch was used.

Scoring: For each correct judgment, a credit of 5 per cent was given. For every error of a quarter of an inch, 5 per cent was deducted from the score. For every error of more than one-quarter of an inch, one per cent. for each half-inch of error was deducted from the score.

VI. Sorting Test.

Materials: A hundred strips of stiff card-board were cut to correspond with twenty of the most commonly used feathers, similar to those employed in the previous test. Five strips of each size were cut.

Instructions: These strips were laid before the subject in a random heap. The subject was told to treat the pieces of card-board as if they were feathers and sort them into piles, according to size. "Lay those that are of the same length and width together. Work as quickly and as carefully as you can."

Scoring: Score the time required for the entire sorting. Credit 100 per cent. for five distinct piles of cards of like size. For every error add five seconds to the subject's time-score. If more than five piles are made, add five seconds to the total time-score for each additional pile. Subjects are graded on the basis of their final time-score.

VI...3. Card Sorting.

Materials: Two packs of ordinary playing cards from which the two-spots had been removed, as 100 cards are scored more easily than 108 cards.

Instructions: The subject was told to sort the pack of cards into four neat piles according to suits. She was shown a sample card of each suit to preclude any errors on the ground of lack of knowledge of the four suits. If a subject is unfamiliar with the cards note should be made of it. All of these subjects knew the suits and played cards "occasionally."

Each subject was directed to sort the cards as rapidly as possible without making any errors.

Scoring: Time was recorded on the individual record blank under Test VI...3. The majority of errors were corrected by the subject who penalized herself because of the extra time spent in rectifying mistakes. The Munsterberg method of penalties was used. (Ref. 46). The subjects were arranged as in the other tests in rank order of merit.

VI. 4. Discrimination of Color Test.

Materials: From a Wholesale Worsted Firm, a quantity of wools, consisting of defective portions of skeins of all colors and grades, was purchased. There were ten shades each of four different colors, Blue, Green, Purple and Orange. Pieces of these shades were cut into fifteen inch lengths and tied in bows. There were forty other shades and colors which were cut also into fifteen inch lengths and made into bows. Four sets were arranged in each of which were placed a complete set (i. e. ten shades), of either the Blue, Green, Purple or Orange Wools. Mixed with each of these four sets were put seventy-five bits of other worsteds, chosen at random. Each set contained exactly the same number of pieces of worsted. Each of these complete sets was put into a large envelope on the outside of which was written the name of the principal color contained therein. i.e. Blue, Green, Orange or Purple. Before each subject a sheet of 8x10 paper was laid.

Instructions: The directions were given orally, as follows: Each of you has been given a sheet of paper. Write your name on top of it. You have been given an envelope that contains bits of worsted of different colors. Those of you who have an envelope upon which Green is written, are to sort out the bits of worsted that are green; in the same way, those of you who have an envelope bearing the word "Blue" are to sort out all the blue shades of wool, and so on with

the Orange and the Purple. Pay no attention to any color save the one you have been told to select. When the signal "Ready" is given, empty the contents of your envelope on the table and begin at once to sort out your color. Put the worsteds that you have been told to select on the white paper before you. Sort the colors as quickly and carefully as you can.

One minute was allowed for the test.

After a girl had sorted one set, she was given another set until each girl had sorted each set.

Scoring: The results of the four color sortings were averaged. There were 85 bits of worsted in each envelope, ten of which were removed from the rest. A credit of 1 was given for each selection of the designated shade, and a deduction of 1 when a color was misplaced. The time was a constant element in all instances, so that the rank order of merit was derived on the basis of accuracy alone.

VI 5. The Motor Control Test.

Materials: On a sheet of 8x11 on legal paper, make lines on the typewriter one-quarter of inch long, approximating thirty-two rows with thirty-eight slants in a row, thus: / / / / /

Instructions: The subject is given a sheet and told at a signal of "Ready" to draw a line through the center of each slant. The Experimenter illustrates. The time allowed for the test is optional with the Experimenter. As this test is intended to determine whether a subject has sufficient stability to persist with steadiness and motor control at an interesting and monotonous task, it is desirable to have the sheet completed, or a time-limit of seven minutes and thirty seconds given. The thirty second periods should be checked off by the subject at the command of the Experimenter, who tells the subject to "check" at 30-second intervals. The Experimenter illustrates by showing the subject a check-mark and she is told to continue until directed to "stop."

Scoring: A standard cross-stroke of approximately an eighth of an inch longer than the obliques was

chosen and only lines which crossed the slant were counted. The number of passable strokes were estimated on this basis. The sheets were graded by three judges who considered them from the standpoint of neatness and precision of stroke, and arranged them first in five piles —Excellent, good, fair, poor and bad. The Experimenter then numbered these five piles 1—2—3—4—5. The three judges then arranged the papers in rank order. The average position obtained by each individual in the Judges' Grading was calculated. The scoring was on the basis of the first grading 90 per cent. and over for the Excellent pile; 80 per cent. to 90 per cent. for the good; 65 per cent. to 80 per cent. for the fair; 30 per cent. to 65 per cent. for the "poor" and 30 per cent. and below for the very poor or "bad" papers.

BIBLIOGRAPHY.

1. H. L. Hollingworth.....Vocational Psychology. Pp. 56 to 79. Chapters V., VII., XI.
2. J. Mc.K. Cattell and L. Farrand.....Physical and Mental Measurements of the Students of Columbia University: Ps. R. 3: 1906: pp. 618 to 648.
3. C. Wissler: The Correlation of Mental and Physical Tests: Ps. Mon. 3; No. 6.
4. H. Munsterberg: Psychology and Industrial Efficiency.
5. E. L. Thorndike: The Psychological Methods of Testing Intelligence. P. 210.
   1. Ed. Ps. Vol. III.
   2. An Introduction to the Theory of Mental & Social Measurements.
6. G. M. Whipple: Mental and Physical Tests: Opposites Test: P. 79: Analogies Tests P. 89. Substitution Test: P. 133. Rote Memory P. 160.
7. Vocational Bureau of Boston, Reports
8. Article in London Times, October, 1920.
9. Army Tests: The Personnel System of the Army. Pub. by War Department.
10. Trade Tests: J. Crosby Chapman.
11. W. D. Scott, Advertising and Selling: Oct. 25, 1915.
12. Report of American Tobacco Co., 1916., edited by J. M. Bruce.
13. H. L. Gardner, The Selection Problem of Cheney Bros., U. S. Bur. Labor Statistics, Bul. 227.
14. R. Feiss: Personal Relationship as a basis of Scientific Management. An. Am. Acad. May, 1915.
15. Dallas Consolidated Railway Co. Reports.
16. Curtis Pub. Co.
17. Report of the Bell Telephone Co. 1921.
18. Roy W. Kelly, Hiring the Worker.
19. St. Elmo Lewis, Industrial Management, March 1917.
20. W. F. Kemble, Choosing Employees by Test.
21. Henry C. Link, Employment Psychology.
22. R. A. Spaeth, Industrial Management April 13, 1920.
23. M. Bloomfield, Employment Management, Selected Articles, H. W. Wilson Co. N. Y. City. Employment Problems Ind. Man. August, 1917.
24. F. F. Taylor, Metropolitan Life Ins. Co., Ind. Man. May 1917. P. 873.
25. M. Jaques, Choosing Employees by Test.
26. K. Murdoch, A Sewing Scale.
27. Nat. Ass'n Corp. Schools. Report of Committee on Vocational Guidance, June 1915.
28. Helen T. Wooley, Mental & Physical Meas. of Working Children Mon. 18: 14. June 1915.
29. Bernard Muscio, Lectures on Industrial Psychology.
30. Brewer and Kelly: A Critical Bibliography of Vocational Guidance, 1917. Harvard Press.
31. Vocational Tests, Teachers College.
32. J. Weidensall: The Mentality of the Criminal Woman., Warwick and York. Pp. 142—158. 176.
33. H. Goddard-Binet Measuring Scale of Intelligence: What it is and How it is to be Used. Tr. Sc. 1911.
34. K. B. Davis., Commercialized Prostitution in N. Y. City. Bur. of Social Hygiene.

35. Report of Dr. Olga Bridgman and Dr. L. Morrow, State School for Girls, Geneva, Ill. "Training Sch."; May, 1912.
36. School and Society: June 23, 1917. P. 747.
37. Mental and Sensory Tests: Idaho Industrial School Report.
38. T. H. Haines: The Measurement of Delinquents. Exper. Ps. 1., 1916.
39. E. L. Thorndike: Educational Psy. Vol. 2. P. 260.
40. W. Stern: The Psychological Methods of Testing Intelligence. Ed. Ps. 13., P. 414.
41. A. I. Gates: The Mnemonic Span for Visual and Auditory Digits. J. Exper. Ps. 1919., I. 393-403.
42-43-44. Psychological Mon., No. 13., Pp. 53-55; 75; 85.
44a. Kirkspatrick: Development and Learning: Archives of Ps. 1909., P. 36.
45. Dr. Augusta Bronner: A Comparative Study of the Delinquent Girl.
46. Munsterberg Method of Penalties: Whipple: Mental and Physical Measurement. Page 313.
47. J. Mc.Keen Cattell and L. Farrand: Ps. R. No. 3: 3. 1896., 618-648.
48. W. G. Chambers: Individual Differences in Grammar Grade Children J. Ed. Ps., 1910; Pp. 61-75.
49. J. E. Wallin: Experimental Oral Euthenics: Dental Cosmos: April, May, 1912. Page 323.
50. E. A. Doll: The A-Test with the Feeble-minded: Train Sch. 10; 13. Pp. 49-57.
51. C. Burt: Experimental Tests: Higher Mental Processes and their Relation to General Intelligence. J. E. Pd. 1: (93-112).
52. S. Wyatt: The Quantitative Investigation of Higher Mental Processes. Br. J. Ps. Vol. 6., 1913. Pp. 109-133.
53. E. L. Thorndike: School and Society. Vo; 9., P. 192.
54. Woodworth and Wells: Ps. Mon. 54., 57.
55. E. L. Thorndike: School and Society. Aug. 12, 1916. P. 261.
56. C. Ritter: Ermudungemessungen: Z. Ps. 24: 1900: 401-404.
57. E. L. Thorndike: Educational Psy. Vol. III. P. 370.
58. B. R. Simpson: Correlation of Mental Abilities. Col. Con. Ed. 53; 1912.
59. Stern, Wm. The Psychological Methods of Testing Intelligence Educational Mono. 1914.
60. Burt. Experimental Tests of General Intelligence Brit. Journal Psy. Vol. 3—P. 94.

F. M. RAPP, Printer,
165 Chambers St., N. Y.

A STUDY OF THE RELATION OF ACCURACY TO SPEED

BY

HENRY E. GARRETT, Ph. D.
(Assistant in Psychology, Columbia)

ARCHIVES OF PSYCHOLOGY
EDITED BY R. S. WOODWORTH
No. 56

NEW YORK
MAY, 1922

AGENTS: G. E. STECHERT & CO. : London (2 Star Yard, Carey St., W. C.); Paris (16, rue de Condé)

ACKNOWLEDGMENTS

The writer is indebted to Prof. A. T. Poffenberger for suggesting the problem of this paper, and for his constant interest and criticism. For many suggestions as to method of treatment and conduct of the experiment he is indebted to Prof. R. S. Woodworth, and Prof. H. L. Hollingworth. He also wishes to express his appreciation to the six men who faithfully acted as subjects throughout a long and rather tedious series of experiments.

TABLE OF CONTENTS

INTRODUCTION 5

SECTION I. HISTORICAL SURVEY 8

SECTION II. EXPERIMENTS ON SPEED AND ACCURACY IN THE PERCEPTION OF SMALL DIFFERENCES 24

1.—Lifted Weights

2.—Linear Magnitudes

3.—Handwriting Specimens

SECTION III. EXPERIMENTS ON SPEED AND ACCURACY OF COORDINATION 80

1.—Thrusting Experiment

2.—Tracing Experiment

SECTION IV. RECAPITULATION AND SUMMARY 100

A Study of the Relation of Accuracy to Speed

INTRODUCTION

The question of the relation of the speed or the quickness of a judgment, perception, or movement to its accuracy is one of no little importance. The desirability of accuracy has never been questioned; but speed is so variable and so individual a matter that it is rather difficult to think of a general relationship between the two factors. In fact, our daily experience reveals a good deal of variation in the time necessary for judgment under different conditions. There would, very probably, be considerable difference in both the character and the exactness of a judgment rendered after one second, and one given after one minute. Given a situation which remained essentially constant, and we should expect that if adaptation and fatigue were eliminated, the judgment or the perception would grow in accuracy with the increase in time taken to make it. The factors involved would gradually coalesce and in consequence, there would be an increase in the confidence with which the judgment was made. No doubt this is what actually does happen in many sensory judgments and perceptions. Oftentimes the perception of difference or similarity increases to a point of maximum clearness, beyond which a prolongation of time results only in doubt, or hesitations:—instead of the decision increasing in accuracy and certainty, the reverse is true, and inaccuracy and uncertainty arise. This suggests an optimal time for perception of difference, which varies with the individual, stimuli, and conditions. Incorrect judgments would then be those given before or after this optimal point had been reached.

In the case of movements, simple and complex, the problem, being more objective, is more easily investigated; though the

information here is still far from exact. Everyday knowledge seems to indicate that, in general, accuracy diminishes as speed increases, but there is little detailed information beyond the bare statement. We do not know, for example, except in certain cases[1] which have been studied, whether accuracy falls off regularly as the speed increases or whether there is a point at which accuracy is greater than at higher or lower rates; nor do we know how the increases in speed affect accuracy.

The experiments described in the following paper are concerned with certain aspects of the speed-accuracy relation; more exactly with this relation as it shows itself in the perception of small differences or in comparison of stimuli close together on the quantitative scale. In the present experiments, an attempt has been made to answer the following questions:—

1.—Will accuracy fall off regularly as speed increases, or is there an "optimal" point, as suggested, at which the accuracy is higher than at lower or higher rates?

2.—How do the increases in speed affect the accuracy, e.g. proportionally, or otherwise?

3.—Do speed and accuracy have essentially the same relation in the perception of differences, as in simple motor coordinations, or does the accuracy behave differently (with increases in speed) in the two cases?

4.—Does the same individual hold his position (relative) in the different situations,—is he fast and accurate, slow and accurate, etc., or can individuals be so classified?

As material for the experiments on the judgments of differences, I used lifted weights, lines, and specimens of handwriting. In the lifted weights and the handwriting, the stimuli were presented successively, the standard always preceding the variable stimulus; with the lines the stimuli were presented simultaneously. In every case the speed, the period of exposure or judgment or both, was the controlled factor and the accuracy was measured by the number of correct responses. For comparison with the results obtained from this group of experiments, two fairly simple tests of coordination were employed, in order to see whether the relationship of speed and accuracy found in the perception experiments also obtained in the field of movement.

[1]Woodworth—Accuracy of Voluntary Movement, PsM, 1899-1900, Vol. 3. XIII.

A brief resume of the studies made of the speed-accuracy relation precedes the experiments. In this review, the studies have been classified generally in the same order in which the experiments are given—weights, lines, writing, coordination. All relevant studies, however, no matter what the experimental material employed, have been included.

SECTION 1

HISTORICAL SURVEY

1. Lifted Weights

The object of the greater part of the experimental work done on Lifted Weights has been the testing of the validity of Weber's Law, the measurement of sensitivity, the analysis of the factors which enter into a sensory judgment, etc. The question of the relation of the time or rate of lift, or the time given for judgment, to the accuracy of discrimination has usually arisen incidentally as the result of a variation in procedure, or as an interesting sidelight on the main problem:—in no instance has it been the chief interest. I have, therefore, in citing results from the classical work done in this field, restricted myself to those references which are directly pertinent to the problem dealt with in this paper. For, as a matter of fact, to review, even briefly, the extensive literature which has grown up around the subject of Lifted Weights would be impossible in a limited space, and for the present purpose, it would be unnecessary.

In the year 1858, Fechner[2] in an experiment on lifted weights with himself as subject made 16,384 lifts using a stanadard of 1000 gms. and two comparison weights differing from the standard by 40 gms. and 80 gms. respectively. Four time intervals for lifting were used: ½″, 1″, 2″, 4″,—and the lifts were divided equally between the left and the right hands. In the 4″ interval, fatigue clearly entered as a factor; nevertheless, the accuracy of judgment varied but very slightly for all rates, apparently the only real effect of the long interval[3] being an increase in the time error: the tendency to overestimate the second weight. This caused Fechner to conclude that slow lifting results in the overestimation of the second weight. He states further[4] that large differences between weights are most easily perceived by rapid lifting, small differences by slow lifting. Fatigue[5] tends to increase the general acuity of perception; which phenomenon, Fechner explains as due

[2]Elemente der Psychophysik, Erster Theil, Leipzig, 1889, p. 305ff.
[3]Op. cit. p. 307.
[4]Revision, 364.
[5]Op. cit. 307-323.

to the more active circulation, as shown, for example, in the increased pulse rate. This increased flow of blood renders the brain more susceptible to fine differences, and results in a more acute sensitivity.

After Fechner, the most elaborate researches on lifted weights are those of G. E. Müller and his pupils Martin, Steffens, and Schumann.[6] These investigators were interested chiefly in the analysis and description of the psychological factors which enter into our judgments of lifted weights. One of their distinctive contributions is the idea of the motor "Einstellung," or "set" of the muscles for certain stimuli, eg. light or heavy weights. For example, a person who has been lifting weights in the order light-heavy for some time, on substituting a lighter weight for the heavy, will find that the 2nd weight seems much lighter than it actually is. This is explained as due to a persistence of set or adjustment of the muscles for the first weight combination, and the influence of this set on the second weight pair. The experiment of Steffens[7] supports this conception of the persistence of set. Motor set is also held to be one cause of constant errors, since weights lifted with great rapidity tend to be underestimated, and those lifted slowly overestimated.

In line with the motor factor is the discovery of Martin and Müller[8] of the dependence of judgment in the case of lifted weights on the "absolute impression" of the standard or variable weight, present to consciousness in tactual or kinaesthetic terms—rather than on the memory image or on an actual comparison of the two weights. Other more theoretical influences in determining judgment found by these authors are the "type tendency of judgment"[9] and the "general tendency of judgment"[10] that the percent of right cases is greater when the standard preceeds than when it follows the variable. The question of the "absolute impression," and the judgment

[6]Martin & Müller: Zur Analyse der Unterschiedsempfindlichkeit, Liepzig, 1899. Steffens, Laura: Über die motorische Einstellung,—Zeitschrift für Psychologie und Physiologie der Sinnersorgane, 1900, 23, 240-308. Müller & Schumann: Über die psychologischen Grundlagen der Vergleichung gehobener Gewichte, Archiv für die Gesamte Physiologie, 1899, *14*, 87-112.

[7]Op. cit. 274.
[8]Op. cit. 44ff.
[9]Op. cit. 29ff.
[10]Op. cit. 64ff.

tendency are considered at some length in the present experiment (p. 38). Martin and Müller[11] determined the time required for a judgment by counting the metronome beats without the subject's knowledge. They find that the discrimination of differences requires a shorter time than the recognition of equality, and that the greater the difference between the two weights the shorter the judgment time.

The fact that a weight seems lighter when lifted rapidly than when lifted slowly, lead Muller and Schumann[12] to conclude that the basis of judgment is the rapidity with which a weight is lifted. This is the Müller-Schumann Theory: a light weight is lifted higher and faster than a heavy weight and hence an O's decision depends on the ease with which a weight's inertia can be overcome.

Fullerton and Cattell[13] supplement a research on lifted weights with a series of experiments, in one of which (p. 131) the rate of movement was varied so that one weight was lifted four times as rapidly as the other "either by being lifted higher in the same time, or the same distance more quickly." The height of lift was, for weights one and two respectively, 16 cm. 4 cm., or 4 cm.-16 cm.; and in the experiment in which the time was varied, the time of lift was 2″-½″, or ½″-2″. Two weights, a standard of 100 gms. and a comparison of 108 gms. were used; and 200 experiments of each sort were made on two subjects. The failure of the PE's to change appreciably, cause Fullerton and Cattell to conclude (in contradiction to the theory of Müller and Schumann) that we do not judge differences in weights by the rate with which they are lifted; rather, the chief factors in the comparison they believe to be cutaneous sensations in the fingers against the weights, eg. to prevent them from slipping, and muscular sensations in the arm and fingers. Further confirmation of this fact, according to these authors, is to be found in the fact that we judge the force of a movement better than the time or the distance through which the movement is made.

The work of Jacobi[14] has a bearing on the general problem of the influence of the speed of the lift on accuracy. With the

[11] Op. cit. 196.

[12] See Zur Analyse der Unterschiedsempfindlichkeit, 207-208.

[13] On the Perception of Small Differences, 1892, 117-132.

[14] Untersuchungen über den Muskelsinn, Archiv. für Exp. Path. und Pharmakologie, 1893, *32*, 49-100.

object of seeing if the heavier of two weights requires more time to "get started" than the light, and whether this delay or inertia is not the factor that determines judgment, Jacobi used an apparatus which registered the moment at which the weights began to rise. He found that when the difference in time between the two weights was less than .08″ the weights were called equal; from .08″-.12″ the weights were often called equal; but when the difference in time was greater than .12″ the weights lifted more slowly was always judged the heavier. The conclusion drawn by Jacobi was that weights are compared and judged in terms of the time necessary to overcome their inertia. Jacobi states further that if we surmise in advance that one of two weights is the heavier, we lift it wth greater force, and if the two weights take about the same time in the lifting, we judge our surmise to have been correct.

The work of Jacobi, while suggestive, can not be accepted strictly at face value: Woodworth[15] in commenting on his results, remarks that his experiments were not numerous enough to establish a correlation between time and judgment.

Claparede's work[16] on the size-weight illusion, seems to show that both inertia and height of lift are factors in the comparison of weights. Claparede used three cubes 512, 1728, 4096 cc. in volume respectively, each weighing 345 gms. Each weight was raised by a ring attached to its upper face, and an electric contact was made when lifting began. The height to which the weights were lifted was also recorded. On the average, it took .12″ to overcome the inertia of the largest weight, .21″ the medium weight, and .62″ the smallest weight. The largest weight was raised 25 mm., the medium 20 mm., and the smallest 10 mm. Since the illusion causes the largest weight to be judged the lightest, there is seen to be a correlation between the speed with which inertia is overcome, and the judgment of heavier or lighter: and also a correlation between the height to which weights are raised and judgment.

Kinnaman[17] finds the basis of judgment to be the "memory of a former change of sensation as compared with present

[15] Le Mouvement, Paris, 1903, 137.

[16] La Vitesse des soulevements lors des illusions de poids, L'Année Psychologique, 1901, *7*, 646-648.

[17] A Comparison of Judgments for Weights Lifted with the Hand and the Foot, Amer. Jour. Psy. 1900, *12*, 240-263.

changing sensation."[18] In lifting a weight there is probably a normal rate of change in sensation at which the total amount of change can be judged most accurately, varying with different observers. The speed, according to Kinnaman, may be either too rapid or too slow for taking these changes into consciousness, and in this way affect the judgment.

In some of the more recent work,—Urban,[19] Fernberger,[20] and Brown,[21] the relation of the time of judgment to accuracy has not been an object of study, and the rate of lift has been definitely regulated, or kept fairly constant throughout. Urban and Fernberger used a metronome set to beat at 92 times per minute, with bell marking the fourth beat. Four beats were given each weight, so that each lift occupied a little more than 2½". In Brown's experiments, the speed and the height of lift, though left to the subject, were fairly constant; each lift occupied about ¾", with an interval of ¼", between lifts, and an interval of about 3¼" between weight pairs. As Brown used only one subject it was fairly easy to keep the time relations standard.

2. Linear Magnitudes

In most of the experiments with linear magnitudes, the discrimination time for differences has been under the subject's control, and has not been arbitrarily varied as in the experiment described in this paper. In the method usually employed, the stimuli, eg. two lines differing in length, are exposed, the subject reacts to the longer or the shorter (according to instructions) by releasing the appropriate key, and the time of reaction is registered on a chronoscope. The accuracy is fixed, therefore, and the time of response is the measured factor in the present experiment, on the other hand, the TIME is fixed (or varied only by the experimenter) and the accuracy of perception difference is the measured factor.

In the following brief historical summary, I have attempted to review only those investigations which have a bearing on the general problem of the relation of time of perception to accuracy in the case of linear magnitudes. Only those results have been included which are relevant to the specific problem of the relation of Accuracy to Speed.

[18] Op. cit. 257.
[19] The Application of Stat. Methods to Problems of Psychophysics, 1908.
[20] On the Relation of the Methods of Just Perceptible Differences and Constant Stimuli, Psy. Rev. Monog., 1913. *14*, No. 61.
[21] The Judgment of Difference, Univ. of California Pub. 1910, Vol. 1.

One of the first experiments to bring out clearly the effect of the time factor on the perception of differences between linear magnitudes, is that of Münsterburg[22] in which the object was to test the psychophysic law, eg. "that equal subjective differences are correlated with equal objective relations of stimuli." Münsterburg used the chain method of reaction. One person reacted and gave stimulus to the second, the second to the third and so on until the last person gave the stimulus again to the first. The average time of discrimination was found by subtracting the reaction time of the first person from the total time elapsing between first stimulus and the last reaction and dividing the result by the number of subjects: in this case five. The stimuli, which were exposed in a rather complicated apparatus, were lines of 2.5-5-7.5, 4-5-6, 4.5-5-5.5 mm., respectively, with four multiples of each: making three main groups in all. A disc containing a line equal in length to the second line of each group was placed before the observer. Three fingers of the right hand rested on keys, which were pressed according to the stimulus exposed: the first, if the line exposed were shorter than the standard, the second if equal, and the third if longer. Münsterburg's results indicate that, in general, the psycho-physical law holds, eg. that it requires as long to discriminate between 5 and 2.5 mm., as between 10 and 5, or 15 and 7.5 mm. He notes, however, a "perfectly regular variation from the law" in that the difficulty of discrimination decreases slightly but regularly with increasing length of line. It requires for example, .1 sec. less to discriminate between 27 and 30 mm. (682 sigma) than between 4.5 and 5 mm. (792 sigma) lines one-sixth as long. Münsterburg qualifies his conclusion, therefore, by stating that "the stronger effect of the relative differences of stimuli is constantly influenced by the weaker effect of the absolute differences of the stimuli."

It is worth noting that Münsterburg worked throughout with lines, the differences between which were easily perceptible; and that the only direct bearing which his work has on the Speed-Accuracy relation is the discovery that once having found the time necessary for perceiving the difference between say, 20 and 30 mm. we should expect the time for accurate discrimination of 10 and 15 mm. to be equally as long.

[22] A Psychometric Investigation of the Psychophysic Law, Psy. Rev., 1894, *1*, 45-51.

In a series of earlier experiments, Münsterburg[23] examines the conditions which affect the perception of linear magnitudes, paying special attention to the part played by eye movements in visual judgments. The apparatus used was a wooden frame 600 mm. long and 500 mm. wide covered by a green cloth. Both vertical and horizontal lines were exposed on this screen at a distance of 600 mm. from the subject, with and without eye movements, simultaneously, successively, with varying intervals, etc. Both monocular and binocular vision were used and 20,000 experiments were made with 20 lengths, ranging from 10 mm. to 200 mm. in steps of 10 mm. The main results which are of interest to the present problem are given:

1.—Every change of eye movement, fixation or use of the eyes affects our estimations of lengths; muscular sensations or their after images are, therefore, necessary to explain errors.

2.—Distances to the left are usually overestimated, to the right underestimated, due probably to habits formed in reading and writing from left to right.

3.—Distances reproduced after varying intervals are usually overestimated. The variable error increases with increases of interval from 1″-10″.

4.—The variable line as compared with the standard is usually overestimated.

5.—In the estimation of simultaneously presented lengths when the eyes are free to move the AE varies between 1.1% and 2.3%; without eye movements the AE's vary from 3.7% to 4.9%. This is taken as proof of the dependence of our estimations of magnitudes on the intensity of muscular sensation.

Henmon[24] reports three groups of experiments on colors, lines and tones, by the "discrimination time method"[25] the object of which was to measure the difference in sensation by the time it takes to perceive them. Henmon divides his work on lines into two parts: one series of experiments dealing with absolute, and another with relative differences. The lines, which were 2 mm. wide and 10 mm. apart, were drawn with great care: two lines, a standard and a comparison to each card. In the first half of the work, six pairs of lines with

[23] Beiträge zur experimentellen Psychologie, 1889, 125-181.

[24] Time of Perception as a Measure of Differences in Sensation, Archiv. Phil. Psy. and Sc. Methods, 1906.

[25] Ladd & Woodworth, Physiological Psy., 1911, 360.

absolute differences of .5, 1, 1.5, 2, 2.5, 3mm. from a standard of 10 mm. were used; while the material for the second half consisted of standard lines of 5, 10, 15, 20 mm. and variables which differed by 20% from each standard line respectively. The lines were exposed in an exposure apparatus, exposure continuing until the subject reacted by releasing a key with the right or left hand, according as the stimulus to which he was to react appeared on the left or on the right. The time of discrimination was recorded by a Hipp chronoscope. The following conclusions are drawn from Henmon's work on lines.

1.—As the differences in length of lines decrease in arithmetical progression, differences in the time of perception increase in geometrical progression for the region tested.

2.—For equal relative differences the time of perception tends to remain constant, indicating the approximate validity of Weber's Law for the range used.

In his experiments on color, Henmon found that as the difference between a pair of Red and Red-Orange stimuli is lessened, the time of perception is increased, though the time differences are not proportional to the stimuli-differences. With tones varying in pitch by 4, 8, 12, 16 vib. the same general result was obtained: the time of perception increases as the differences between the stimuli decreases, (with considerable individual difference).

Henmon[26] describes a series of experiments with lines, the object of which was to find (1) the relation of the time of judgment to accuracy, (2) to study individual differences in judgment times. The specific problem was to discriminate between two lines, 20 and 20.3 mm. long, arranged horizontally on a card 10 mm. apart. The lines were exposed after the method employed in the previous experiment, exposure continuing until a judgment was made, with the subject reacting one-half the time to the longer, and one-half the time to the shorter line. Four degrees of confidence were recognized; a, perfectly sure, b, fairly sure, c, uncertain, and d, guess. There were three subjects, from each of whom 1000 reactions were taken. Henmon's observations are as follows:

1.—The average times of R-judgments are shorter than for wrong. Wrong judgments are absolutely and relatively more variable.

[26]Time and Accuracy of Judgment, Psy. Rev., 1911, *18*.

2.—Correct judgment of difference seems to grow to a point of maximum clearness; which suggests an optional time varying with the individual, stimulus, and conditions.

3.—There is a high correlation between confidence and accuracy.

4.—Introspection gives evidence that a prolongation of the judgment-time introduces an element of doubt.

5.—There is a uniform decrease in R-cases with decrease in confidence: also an uniform increase in time with decrease in confidence.

Hermon's subjects showed marked individual differences in accuracy, time of perception, and degree of confidence. Accuracy ranged from 68.6% to 83.2%; time of reaction from 304 sigma to 1095 sigma. The degrees of confidence have different values for different individuals, one subject giving "a" 90.4% of the time, and another only 9.3% of the time. Quick subjects were neither more nor less confident than slow. Henmon concludes that the relation of subjective confidence to accuracy is an individual matter.

Simpson[27] among other tests, used two on discrimination of linear magnitudes. In the first, the subject estimated the comparative length of pairs of horizontal lines: 108-100 mm., 106-100 mm., 104-100 mm., 102-100 mm., the lines drawn side by side one pair to a card. In the other test, the subject was required to draw lines equal to a given line, each line being covered as soon as made, so that the next would be an estimate of the standard line and not a copy of the line just drawn. Simpson found that the fast workers made the best records,—that little was gained by hesitation.

Peterson, G. S.[28] reports an experiment with linear magnitudes "undertaken to show the relation between time and amount of difference, time and accuracy, and between accuracy and the amount of difference in two ways: (1) By varying the size of the difference about any standard, (2) By varying the standard with the difference constant."

Peterson used twelve lines, in pairs, arranged about three standard lines 5, 10, 20 mm. long respectively. Each pair of lines consisted of a standard and a variable: four variables

[27]Correlation of Mental Abilities, Columbia Univ. Contributions to Education, Teachers College Series, 1912, *53*.

[28]On the Speed and Accuracy of Judgment of Horizontal Lines, M.A. Essay, Columbia University, 1916.

for each standard being used. The differences between the standard and the variable varied from .5 mm. to 3 mm. For the 5 mm. standard the variables were 5.5 mm., 6 mm., 7 mm., and 8 mm.; for the 10 mm. standard, 9.5, 9, 8, 6 mm.; for the 20 mm. standard 19.5, 19, 18, 16 mm. These lines were arranged one pair to a card, and were placed horizontally in the center of the card, separated by 1 cm. The cards were exposed in a tachistoscope, the subject reacting to the longer line as soon as the difference was perceived. Reaction time was registered on the Hipp chronoscope. On the basis of 4259 reactions taken from three subpects, Peterson concludes that (1) differences in perception time decrease approximately in GP as the difference between lines increase in AP; (2) error times are constantly less than correct times; (3) the more difficult the judgment, the more variable the perception time. Increase in difficulty leads to increase in time to be accurate, and results in fewer R-judgments; while decrease in difficulty leads to a decrease in time to be accurate and results in more R-judgments.

In addition to the results quoted from experiments with weights and lines, there have been several studies made of the speed-accuracy relation in experiments employing other material and apparatus. A brief account of these results will be given.

Angell and Harwood[29] and F. Angell[30] in an experiment with tone discrimination varied the interval between the first or standard tone and the comparison tone from 1″ to 60″. The comparison tone differed from the first by 0, ± 4 or ± 8 vibrations. There were four subjects and more than 6000 judgments were taken, with and without distraction during the interval between stimuli. Angell's results were as follows:

1. There is no loss in accuracy with increase in time interval up to 60″.
2. Distractions have little effect on the accuracy of judgment.
3. Introspections reveal a type of judgment termed "free": judgments delivered with considerable confidence, but without the presence in consciousness of any comparison of standard and variable.

[29]Experiments on Discrimination of Clangs for Different Intervals of Time, Part I, Amer. Jour. Psy. 1899-1900, *11*, 67-79.

[30]Part II, Amer. Jour. Ps. 1901, *12*, 58-79.

Angell F.[31] in a later series of three experiments, studied the influence of time interval on the accuracy of discrimination for different shades of grey. In the first experiment, a disc containing black and white in equal proportions was exposed for 2″ on a color mixer; and after a given interval,—5″, 15″, 30″, 60″—a comparison disc containing 200, 190, 170, 160, degrees of black was shown. Judgment of the comparison in terms of the standard were "darker," "lighter," or "same." The result of this experiment for two subjects indicated that "the accuracy of the judgment is practically independent of the time interval employed." This is true whether the subject makes an effort to "hold fast" to the visual image during interval, or merely remains relaxed with no definite attempt to fix the attention on the stimulus just seen. Two years later in a repetition of the same experiment with two subjects, under more carefully controlled conditions, Angell obtained the same result as to the influence of the time interval on accuracy. In another experiment of the same general sort, and with the same material, Angell measured to tenths of a second the time of the judgment for the four categories, "sure," "fairly sure," "like," i.e. same, and "doubtful." His results showed that "sure" judgments were the shortest, "like" judgments the longest. Angell notes again the presence of "free" judgments: —those in which no direct comparison of the standard and variable seemed to occur.

Whipple, G. M.[32] in an experiment on the discrimination of clangs and tones, varied the interval between stimuli from 2 to 60 sec. Whipple used an Appunn Tonometer; five tones were taken as standards, and the comparison tones differed from the standards by 0, +8, –8, vibrations. The relation of speed of judgment to certainty, and correctness; and the relation of speed of judgment to immediacy and correctness, were considered. It was found that (p. 445-6) immediate judgments were usually both certain and correct, while the delayed or compared judgments were more often uncertain and incorrect.

Bentley[33] in an elaborate study of memory fidelity for bright-

[31] Philos. Studien, 1902, *19*, 1-21.

[32] An Analytic Study of Memory Image and the Process of Judgment in the Discrimination of Clangs and Tones, Amer. Jour. Psy. 1901, *12*, 409-446.

[33] The Memory Image and its Qualitative Fidelity, Amer. Jour. Psy., 1899, *11*, 1-48.

nesses and colors, investigated in the course of his experiment the influence of the time interval between standard and comparison discs on the quality and fidelity of the memory image. Black, white, and colored discs, in various proportions were used as stimuli. Bentley's results showed no loss in accuracy of visual memory up to 6″, and a very slight loss up to 60″. After one minute there was a loss, increasing with the interval.

Woodworth[34] devotes one section of a research on the accuracy of voluntary movement to the relation between the speed of a movement and its accuracy. This last was tested out in two ways, by ruling lines on a drum, and by hitting at a target in various ways. In the first experiment, the requirement was that each line should equal the one immediately preceding it; in the other tests, the subject was required to thrust with a pencil alternately at three dots arranged in the form of an equilateral triangle, and 15 cm. apart respectively, or to hit with a pencil point at the center of each in turn of small squares in a sheet of coordinate paper. Speed was regulated by a metronome. Woodworth found that with increase in the speed of movement, there was a corresponding decrease in accuracy; but that movements at 40 per minute were as accurate as those at 20 per minute, while movements at 140, 160, 180, 200 per minute were all about equally as accurate—or inaccurate. Accuracy, then, did not vary proportionally or regularly with the increases in speed. The analysis of the factors entering into such movements as the ones here described, as well as the effect of variations in interval between movements, adjustment and control of the movement process, etc., have been considered at greater length in the experiment on Thrusting to the present paper. Woodworth suggests that there is an optimum interval between the movements studied, but this point cannot be located with any high degree of certainty as it varies with the right or left hand and the character of the movement.

Morgan, J. J. B.[35] reports an experiment on speed and accuracy of motor adjustments, the object of which was twofold; (1) to find the speed of adjustment to a change in load, (2) to find how accurately one can compensate for changes in

[34] Accuracy of Voluntary Movement, Psy. Rev., Mon. 1899, 27-54.

[35] The Speed and Accuracy of Motor Adjustments, Jour. Exp. Psy., 1917, 2, 225-248.

load by changes in time so as to keep the physical force exerted the same throughout. Morgan's apparatus consisted of a movable carriage, to one end of which was attached a cord to be pulled by the subject; and at the other end, a cord which passed over a movable pulley to which weights were fastened. The carriage ran in grooves hollowed in horizontal guides; a 100 vib. tuning fork attached to it wrote during movement on a flat smoked surface underneath. Morgan found, in the first place, that a subject's idea of his maximum force is largely determined by the resistance to the movement involved; and that the quickness with which adjustment is made indicates that it is of the reflex type. In the second place, it was found that subjects do not possess the ability to adjust the force of a movement accurately, so that the same physical force will be exerted with different weights; in other words, the subjects judged of the force exerted by the time of pull, while they compensated for different loads (when told to exert the same force) by means of a very crude time correction.

Gould[36] in a study of transfer of practice in motor activity, used a test in which both speed and accuracy were required. The general scheme of the test was adapted from McDougall's "Dotting Test."[37] Sheets containing small circles irregularly arranged were fastened on a metal drum, which revolved on a horizontal axis. The subject's task was to hit the middle dot as the rows passed successively by a small window cut in a wooden screen placed before the drum. The drum was set to revolve at four speeds; so that one row of circles passed before the window in 2″, 1.6″, 1.2″, 1″ respectively. Two groups of boys were used as subjects; a group of 61 Prevocational school boys, and a "control" group of 62 academic boys. Both groups performed the experiment twice with the interval of a year between. Using the median score in the first test for the group of PV boys, the accuracy records for the four rates are 114, 115, 111, 105. The loss in accuracy when the speed is doubled is, then, 8%, while the accuracy remains at the same level for the first two rates. After a year, the scores for the same group are 116, 115, 111, 105; the speed, increased by 100%, resulting in a loss in accuracy of 9.5%. For the control group, in the first test the scores are 120, 113, 111, 105, a

[36]Transfer of Manual Accuracy in Prevocational School Boys, M.A. Essay, Columbia, 1917.

[37]Br. Jour. Psy. 1905, 435ff.

loss in accuracy of 12% as speed doubles. After a year the scores of the control group are 114, 111, 108, 107, a loss of about 6% in accuracy. The loss in accuracy was, therefore, not at all proportional to the increase in speed.

Studies in which tests of perception or discrimination, eg. cancellation tests, have been utilized, have a bearing on the problem of the speed accuracy relationship, because correlations have been found for speed and accuracy (in many cases), and because of the time corrections for errors and omissions. These corrections are for the purpose of converting speed and accuracy into one single coefficient.

Cattell and Farrand[38] in scoring the A-test, used with College Freshmen, made a "rough correction" for mistakes by adding to the total time the time that would be required to discriminate and mark the letter omitted or wrongly marked. Cattell and Farrand remark that subjects are slow and accurate, slow and inaccurate, fast and accurate, and fast and inaccurate:—no correlations were found.

Wissler[39] in his A-test with Columbia Freshmen, found a correlation between speed and accuracy of –.28.

Wyatt and Brown[40] in cancellation tests with school children mark +1 for each symbol correctly marked, and –1 for each mistake or omission.

Simpson[41] in the A-test, added 5″ for each A omitted to the total time. In the geometrical form test, he added 3″ to 6″ depending on the symbol omitted. The few errors due to wrong cancellation were neglected.

Whipple[42] reports a correlation for speed and accuracy of –.37, for 50 Grammar School boys, when one lettter was cancelled: and a correlation of -.64 for the same group when four letters were cancelled. With 30 University students, Whipple found a correlation of -.48 when four letters were cancelled.

Woodworth and Wells[43] suggest that in the number checking test, a correction should be made of 2% of the subject's total time as a penalty for each error or omission when one-half of

[38] Psy. Rev. 1896, 3, 641.

[39] The Correlation of Mental and Physical Tests, Psy. Rev. Mon. 1901, p. 20.

[40] Quantitative Investigation of Higher Mental Processes, Brit. Jour. Psy. 1913, *6*, 114.

[41] Op. cit.

[42] Manual of Mental and Physical Tests, Pt. 1, Simpler Processes, 1914, 318.

[43] Association Tests, Psy. Rev. Mon. 1911, *13*.

the blank is used; and a correction of 1% when the whole blank is used.

Strong[44] when using the Woodworth-Wells Number Checking Test, made a correction by finding the time for cancellation of a single symbol and adding twice this time to the recorded time.

Thorndike[45] reports an investigation on the relation of speed and accuracy in addition in which 671 students were used as subjects. The best 65 students averaged 15 additions per 100 sec., the worst 20 averaged 37 in the same time. The best group made 7 errors in 100 additions, the slow group 17½. Thorndike concludes that a person who is rapid will tend to be accurate also.

Bird, Grace E.[46] tested 100 College students with slow and rapid adding of examples taken from the Courtis Tests. The students added for two minutes as rapidly as possible; then for two minutes in which they were cautioned to work slowly and avoid errors. In the first test with the rapid addition, the median number of errors was 3, with a quartile deviation of .5; in the second test with the slow addition, the median number of errors was 4, with a quartile deviation of .8. When introspections were taken, all but three of the subjects mentioned distractions during the slow adding, while only five record any distraction during the rapid adding. The distractions mentioned include a variety of imagery, adding by combining units rather than groups, unnecessary repetition of sums obtained in the process of adding a column, emotional disturbances, physical uneasiness, shifting of attention to the environment, forgetfulness of sum already obtained, losing place, amusement at the experiment, etc. Bird attributes the superiority of the rapid addition to a difference of adjustment. In slow addition distractions come in, a reversion to childhood habits, eg. counting on the fingers, lip movements, vocalization etc.; in the rapid addition, irrelevant matters are crowded out through the inhibitory nature of the work, and the response becomes more automatic.

Due largely to differences in the "setting" of the experiments, methods and materials, it would seem to be rather

[44] A Comparison between Experimental Data and Clinical Results in Manic-Depressive Insanity, Amer. Jour. Psy. 1913, *24*, 83.

[45] Relation of Speed to Accuracy in Addition, Jour. Ed. Psy. 1914, *5*, 536.

[46] The Devious Path of Slow Work, Jour. Ed. Psy. 1922, *13*, 50.

difficult, or impossible, to gain any clear conception of the speed-accuracy relation, in general, from the above resume. It does appear, however, that the time given for perception or discrimination in those experiments, eg. lines, weights, tones, brightnesses, etc., in which differences are to be perceived, has little effect over a fairly wide range of time intervals; while in experiments dealing largely with cancellation the accuracy-speed relation is inverse. In movements requiring accurate and precise coordination, drawing lines, thrusting at dots, etc., the loss in accuracy is very small in comparison with the increases in rate, and in adding, both a motor and mental function, speed and accuracy are directly related. Practically nothing can be said in regard to an optimal interval for perception of difference, as little attention has been given this question; in simple movements, however, the evidence in favor of such an interval is fairly conclusive.

SECTION II.

1. Speed and Accuracy in Lifted Weights.

PROBLEM

In the experiment here described, the problem was to find to what extent the accuracy of weight discrimination is affected by the rate at which the weights are lifted, and the length of the interval between successive lifts. This question may well be resolved into two parts:—

(1) What effect does the rate of lifting have on the accuracy of judgment? and (2) What effect does the time-interval for judgment have on the accuracy of discrimination? In addition to these questions, further interesting corallaries grow out of the main problem. Some of these are noted. As the rate of lifting changes, what change will there be in the so-called threshold? Is the tendency to overestimate the second weight increased or diminished by the increase in the lifting rate? Is there an "optimal rate" at which accuracy is higher than at lower or higher rates? What is the relation between the confidence of the subject in the accuracy of his judgment, and the actual results obtained at the different rates? There arises, too, the question of individual differences in the course of the experiment, and, though it is realized that not much can be generalized about differences among individuals with only six subjects, still a comparison of results is suggestive, and in an experiment of this nature often brings out real differences.

APPARATUS AND PROCEDURE

The Method of Right and Wrong Cases in the simplified form advocated by Jastrow[47] and Fullerton and Cattell[48] was used. This method seemed preferable to the older and more complex form because accuracy percentages, thresholds, C E's etc., can, by means of it, be easily and quickly calculated. For other advantages and reasons for the elimination of the "equal" judgments, see Woodworth, Archiv. of Psy. 1914, No. 30, 65-68.

[47] Amer. Jour. Psy. 1887, *2*, 271-309.

[48] On the Perception of Small Differences, 1892, 119-120.

The weights used were cylindrical wooden boxes 4½ cm. high, and 6½ cm. in diameter. These boxes were loaded with paraffin and shot and were carefully weighed. There were 14 weights in all:—7 standard weights, and 7 comparison weights. The standard weights were all 100 gms.; the comparison weights formed a series from 88 gm. to 112 gm. in steps of 4 gms. The boxes were all varnished black and were indistinguishable in appearance.

The technique employed in lifting the weights was a modified form of Urban's[49] method. The weights were arranged in pairs at regular intervals along the circumference of a table with a revolving top. A screen with an aperture cut for the lifter's hand shut off the subject's view of the weights, which, by means of a turning table could be brought in succession directly under the subject's hand. In this way the space error was eliminated. Any error due to the time order was constant, as the standard was always lifted first. The use of comparison weights which extend below and above the standard would seem, however, to obviate the necessity of using more than one time order.

When the subject had lifted the second weight of any pair, he was required to say whether it was heavier or lighter than the standard just lifted, and to state his confidence in his decision. The letters *a*, *b*, and *c* were used as measures of subjective confidence, *a* meaning "sure" of the correctness of the decision, *b* "fairly sure," and *c* "uncertain" or doubtful.

One experiment consisted of five revolutions of the table:—that is to say, 5 judgments on each weight pair, or 35 judgments in all. Between experiments there was a rest interval of 3 min., during which the experimenter rearranged the weights. The five series of settings used for the comparison weights were as follows:

1.—104, 92, 108, 88, 96, 100, 112.
2.—96, 104, 108, 112, 92, 100, 88.
3.—112, 104, 96, 100, 92, 108, 88.
4.—88, 100, 108, 92, 104, 96, 112.
5.—112, 88, 92, 96, 100, 104, 108.

Since the rate of lifting and the time of judgment were the important factors in this experiment, all the lifting was regulated by a bell metronome. Five rates of lifting were

[49]Applications of Stat. Methods to Problems of Psychophysics, 1906, 1-18.

used, with the metronome set at 60, 96, 120, 160, 240. This last rate was obtained by removing the runner from the pendulum of the metronome. When the beats were then timed with a stop watch, there were approximately 4 to the second. Four beats of the metronome were given for each weight, with a four-beat interval for judgment between successive weight pairs. After some experimentation it was found to be easier to let the bell mark the first movement in the lifting process, i.e. that of grasping the weight. On the bell, therefore, the subject's hand grasped the weight, at the second beat raised it, at the third beat returned it to the table, and at the fourth beat came up again.

The different times for lifting with their judgment intervals are designated as rates 1, 2, 3, 4, 5. At Rate 1, with the metronome at 60, 4 seconds were given each weight with a 4-second interval between weight pairs; at Rate 2, metronome at 96, 2½ seconds were given each weight, with a 2½-second interval; at Rate 3, metronome at 120, 2 seconds were given each weight, with a 2-sec. interval; at Rate 4, metronome at 160, 1½ seconds were given each weight, with a 1½-sec. interval; and at Rate 5, metronome at 240, 1 sec. was given each weight, with a 1-sec. interval. It must be remembered that the number of seconds "given each weight" does not mean the number of seconds that the weight was actually lifted as the procedure described above makes clear: only one-half of the total time was used for lifting.

One variation in method must be noted: namely that judgments were required on the 100 gm. weight when employed as a comparison weight, with a 100 gm. weight as standard. Since "equal" judgments were excluded, the inclusion of a weight pair objectively equal, was, in a sense, a deception and is not strictly in accordance with the simplified form of the Method of Right and Wrong Cases. In fact Jastrow[50] warns against it. My object was, however, to find the proportion of "heavier" and "lighter" judgments on a comparison weight equal to the standard, and to see how this proportion changed as the rate of lifting increased: further than this the results were not used. The fact that the threshold for chance was below 100 gms. at all lifting rates, and that the 100-100 gm. pair were, therefore, not subjectively equal, offers some evi-

[50] "A Critique of Psychophysical Methods," Amer. Jour. Psy. 1888, *1*, p. 288.

dence in justification of the inclusion of the 100 gm. weight as a comparison.

Brief introspections were taken in the course of the experiment as to the criteria of judgment, feeling of confidence, ease with which the various rhythms could be followed; and any hesitation or apparent discomfort on the part of a subject at a given rate was noted by the experimenter, as none of the subjects were trained in careful or elaborate introspection, these records are used chiefly as a commentary on the objective results.

FATIGUE AND PRACTICE

In a series of experiments extending over a fairly long period of time, fatigue must be considered as a factor. Accordingly, in addition to a rest period of 3 minutes between each group of 35 lifts, i.e. 5 revolutions of the table, no series of experiments extended for longer than one hour, so that fatigue could hardly have been present to any appreciable degree. Then, too, the weights were relatively very light. Unfortunately, a definite time of day could not be set, and the experiments were made both morning and afternoon as the subjects were available.

"The consensus of opinion," according to Whipple[51] is "that, at least in comparison with many other mental activities, the discrimination of lifted weights is but little affected by practice." Spearman,[52] hovever, reports that his subjects have "invariably" shown improvement in the first fifteen minutes, and that a preliminary practice series is necessary if reliable records are desired. Fernberger[53] also finds considerable practice at the beginning of an experimental series. He insists that at least 50 determinations should be made on each weight, if a real measure of sensitivity is required; ten determinations are practically useless. To offset the practice factor, each subject in the present experiment was given a

[51]Manual of Mental and Physical Tests, I, 228.

[52]Gen. Intell. Objectively Determined and Measured, Amer. Jour. Psy. 1904, *15*, 233.

[53]The Effects of Practice in its Initial Stages in Lifted Weight Experiments and its bearing on Anthropometric Measurements, A. J. Ps., 1916, Vol. 27, 271.

series of preliminary lifts, to aid him in securing technique, as well as familiarity with the different rhythms. Furthermore, the experiments were taken in groups alternating from one rate to another, so that whatever practice there might be present would be spread over the whole experiment, and not be cumulative at any one rate. Records from the different rates can, therefore, be fairly compared.

RESULTS

There were six men who acted as subjects in this experiment: Mz and Wn, both graduate students in psychology with some laboratory training, and Ad, Dw, Ant, and Sc, all Columbia College undergraduates with at least one year's work in psychology.

Table I, page 32, contains the results of the entire experiment. It is made up of 5 sections, each containing the results of 2100 lifts made by six subjects at a given lifting rate. Each section gives the number of right and wrong cases for each weight difference, the total number of right and wrong cases, the % of right cases, for each observer, and the % of right determinations for each weight. The total % of right cases for a given rate is also given; with the PE's of these %'s calculated from the formula: $PE\ (\text{tr-obt. Av.}) = \frac{P\,E\ (\text{dis.})}{\sqrt{N}}$ These values (PE's) indicate how much—with the chances even—the obtained averages may be expected to differ from the true averages.

Diagram I represents, graphically, each observer's accuracy record, in % of right cases, for each of the 5 rates. The diagram is schematic; the rates are laid off at equal distances along the base line, and the accuracy records in %'s right are placed on the Y-axis. The heavy line is the composite accuracy record for all six observers throughout the 5 experiments.

This general accuracy curve, which, like the curves for the individual subjects is plotted from the data of Table I, shows a slight rise in accuracy as the rate of lifting increases from the data of Table I, shows a slight rise in accuracy as the rate of lifting increases from Rate 1 to Rate 3; then a decrease at Rate 4, and a further and more pronounced "drop" at Rate 5. The following table gives the reliability of these changes in accuracy:

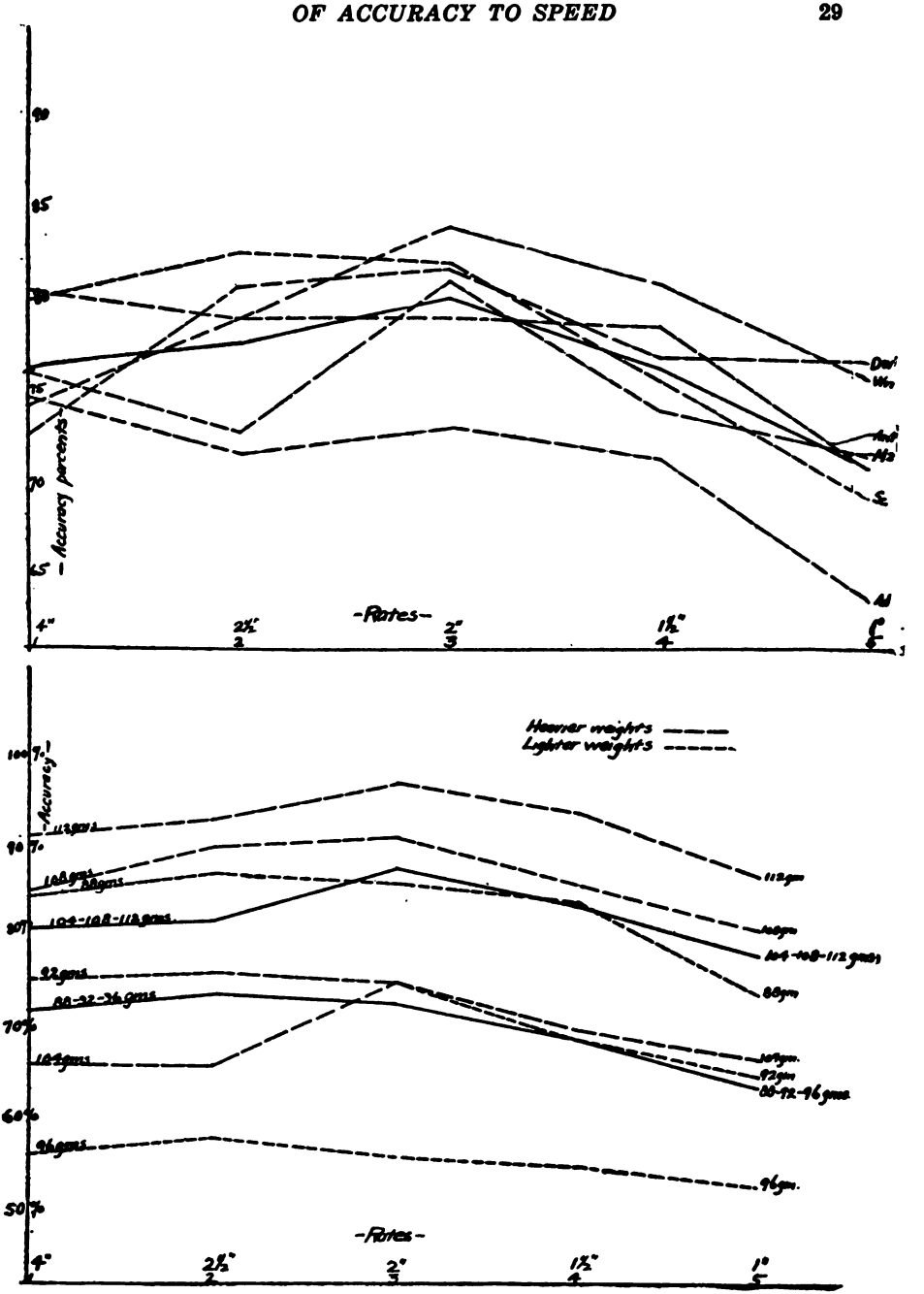


Diagram 1.—Accuracy records, (total % right), for each rate ——————
Individual records of 6 O's .....
Diagram 2.—Accuracy records, for heavier weights (104-108-112) and lighter weights (88-92-96).

Table II

	Rate 1	Rate 2	Rate 3	Rate 4	Rate 5
Aver.	76.3	77.6	80.1	76.3	70.9
PE-aver.	1.51	1.73	1.71	1.64	1.60
Diffs in Av. (fm. R. 3)	3.8	2.5	——	3.8	9.2
P. E.'s diffs.	2.28	2.43	——	2.37	2.29

Probability that true diff. between Rate 3 & 1=0 or less=1301 in 10,000 (Roughly 1 in 8)
" " " " " " 3 & 2=0 or less=2437 in 10,000 (Roughly 1 in 4)
" " " " " " 3 & 4=0 or less=1403 in 10,000 (Roughly 1 in 7)
" " " " " " 3 & 5=0 or less= 35 in 10,000 (Roughly 1 in 250)

The reliability of the differences between the above rates was calculated from the formula:

P. E. (true-obtained diff. A-B $= \sqrt{PE^2 \text{ (true-obt. A)} + PE^2}$ (true-obt. B) in which the averages and the PE (av.) were taken from Table I. In every instance the rates 1, 2, 4, 5, were compared with Rate 3, since the accuracy curve reaches its maximum at this point. The results of these comparisons furnish some interesting data. Rates 3 and 2 are very close together, the chances of a real difference (3 in 4) being too small to indicate any real superiority of Rate 3. The chances of a real difference between Rate 3, and Rates 4 and 1, are 6 in 7 and 7 in 8, large enough to indicate, though not to guarantee, an actual change in accuracy from 1 to 3 to 4. Rate 5 is a hopeless competitor, the chances being about 1 in 250 that it is equal or greater in accuracy than Rate 3. These figures are taken to give fairly conclusive evidence of an improvement in accuracy, from a lifting rate of 4″ with a 4″ interval to a lifting rate of 2″ with a 2″ interval; while with a further increase in lifting rate to 1″ there is good evidence of a considerable loss in ability to discriminate between weights.

A further way of investigating the changes in accuracy with changes in lifting rate, is to study the course of right comparisons for each weight-difference from Rate 1 to Rate 5. This enables us to see how the change in rate affects each comparison weight specifically. Diagram 2 pictures, graphically, the course of the accuracy curves for each comparison weight from the slowest to the fastest rate; as well as the composite record of the 3 weights heavier, and the 3 weights lighter, than the standard weight. The rates occupy equal

distances on the base line, while the %'s of right cases for all six subjects are measured off on the Y-axis. Except for a sudden rise for the 104 gm. weight at Rate 3, the curves of all the comparison weights are regular and of much the same general form. All show the same tendency to rise gradually from Rate 1 up through Rates 2 and 3, and then fall off at Rates 4 and 5.

The curves for all the weights heavier than 100 gms. reach their highest point (maximum accuracy) at Rate 3; while the curves for the three weights lighter than 100 gms. tend to "peak" (in so far as they peak at all) at Rate 2. There is very little change in the course of accuracy for the 92 gm. weight from Rate 1 to Rate 3; and the 96 gm. weight was judged about equally well or poorly at all five rates. The composite records for the heavier and the lighter weights show very clearly this general tendency of the two classes of weights. In the following table the reliabilities of the differences between the accuracy records at the different rates for the heavier and lighter weights, separately, have been calculated after the plan of Table II.

Table III

(88-92-96 gm. weights)

	Rate 1	Rate 2	Rate 3	Rate 4	Rate 5
Aver.	71.90	73.45	72.55	69.11	64.00
PE-aver.	2.08	2.62	2.24	2.31	2.08
Diffs in Av. (fm. R. 2)	1.55	——	.90	4.34	9.45
PE's (dif.)	3.35	——	3.45	3.49	3.35

Probability that true diff. between Rate 2 & 1=0 or less=3781 in 10,000 (Roughly 1 in 3)
" " " " " " 2 & 3=0 or less=4304 in 10,000 (Roughly 1 in 2)
" " " " " " 2 & 4=0 or less=1978 in 10,000 (Roughly 1 in 5)
" " " " " " 2 & 5=0 or less= 287 in 10,000 (Roughly 1 in 35)

(104-108-112)

	Rate 1	Rate 2	Rate 3	Rate 4	Rate 5
Aver.	80.67	81.80	87.67	83.44	78.11
PE-aver.	1.93	2.15	1.93	1.83	1.84
Diffs in Av. (fm. R. 3)	7.0	5.87	——	4.23	9.56
PE's (diff.)	2.73	2.89	——	2.66	2.67

Probability that true diff. between Rate 3 & 1=0 or less= 421 in 10,000 (Roughly 1 in 23)
" " " " " " 3 & 2=0 or less= 857 in 10,000 (Roughly 1 in 12)
" " " " " " 3 & 4=0 or less=1403 in 10,000 (Roughly 1 in 7)
" " " " " " 3 & 5=0 or less= 79 in 10,000 (Roughly 1 in 125)

TABLE I

Discrimination of Lifted Weights:—Method Of Right & Wrong Cases:—Six Subjects:—10,500 lifts:—2100 at each of Five Rates of Lifting, (The 100 gm. weight is not included in the totals)

Rate 1. 4"

	88		92		96		100		2100 Lifts. 104		108		112		Tot.		%R
Sub.	R	W	R	W	R	W	+	−	R	W	R	W	R	W	R	W	R
Mz	36	14	41	9	27	23	23	27	35	15	39	11	50	0	228	72	76.0
Ad	42	8	32	18	29	21	28	22	33	17	42	9	46	4	224	76	74.7
Dw	42	8	38	12	30	20	23	27	29	21	40	10	39	11	218	82	72.7
Wn	43	7	36	14	25	25	30	20	32	18	45	5	42	8	223	77	74.3
Ant	43	7	35	15	28	22	32	18	39	11	48	2	48	2	241	59	80.3
Sc	47	3	43	7	30	20	24	26	30	20	42	8	47	3	239	61	80.0
Sum	253	47	225	75	169	131	160	140	198	102	256	44	272	28	1373	427	
%R		84.3		75.0		56.3		53.3		66.0		85.3		90.7		76.3	
AD		4.3		6.7		3.0				5.3		5.1		6.4		PE(Av)=1.51	

Rate 2. 2½"

	88		92		96		100		2100 Lifts. 104		108		112		Tot.		%R
Sub.	R	W	R	W	R	W	+	−	R	W	R	W	R	W	R	W	R
Mz	42	8	30	20	25	25	34	16	34	16	48	2	39	11	218	82	72.7
Ad	41	9	32	18	33	17	34	16	27	23	39	11	43	7	215	85	71.7
Dw	47	3	45	5	38	12	19	31	30	20	39	11	43	7	242	58	80.7
Wn	45	5	36	14	21	29	32	18	38	12	47	3	50	0	237	63	79.0
Ant	42	8	36	14	23	27	36	14	37	13	49	1	50	0	237	63	79.0
Sc	43	7	48	2	34	16	27	23	33	17	46	4	44	6	248	52	82.7
Sum	260	40	227	73	174	126	182	118	199	101	268	32	269	31	1397	403	
%R		87.0		75.7		58.0		60.7		66.3		89.3		89.7		77.6	
AD		3.7		11.6		12.2				6.3		7.5		6.9		PE(Av.)=1.73	

Rate 3. 2"

	88		92		96		100		2100 Lifts. 104		108		112		Tot.		%R
Sub.	R	W	R	W	R	W	+	−	R	W	R	W	R	W	R	W	R
Mz	41	9	38	12	24	26	36	14	43	7	47	3	50	0	243	57	81.0
Ad	36	14	38	12	33	17	21	29	31	19	37	13	44	6	219	81	73.0
Dw	49	1	39	11	28	22	28	22	34	16	45	5	50	0	245	55	81.7
Wn	44	6	36	14	31	19	34	16	43	7	48	2	50	0	252	48	84.0
Ant	43	7	33	17	23	27	33	17	38	12	50	0	50	0	237	63	79.0
Sc	45	5	42	8	30	20	25	25	35	15	46	4	48	2	246	54	82.0
Sum	258	42	226	74	169	131	177	123	224	76	273	27	292	8	1442	358	
%R		86.0		75.3		56.3		59.0		74.7		91.0		97.3		80.1	
AD		6.0		4.2		6.3		9.7		8.0		6.0		3.6		PE(Av.)=1.71	

Rate 4. 1½" 2100 Lifts.

	88		92		96		100		104		108		112		Tot.		%R
Sub.	R	W	R	W	R	W	+	−	R	W	R	W	R	W	R	W	R
Mz	39	11	34	16	28	22	34	16	32	18	43	7	46	4	222	78	74.0
Ad	39	11	29	21	29	21	39	11	32	18	38	12	47	3	214	86	71.3
Dw	41	9	37	13	34	16	26	24	34	16	41	9	44	6	231	69	77.0
Wn	44	6	41	9	25	25	31	19	37	13	47	3	49	1	243	57	81.0
Ant	46	4	28	22	25	25	38	12	41	9	47	3	49	1	236	64	78.7
Sc	42	8	37	13	24	26	25	25	34	16	42	8	48	2	227	73	75.7
Sum	251	49	206	94	165	135	193	107	210	90	258	42	283	17	1373	427	
%R	83.7		68.7		65.0		64.3		70.0		86.0		94.3		76.3		
AD	4.3		8.0		5.7		9.6		5.3		5.3		3.0		PE(Av.)=1.64		

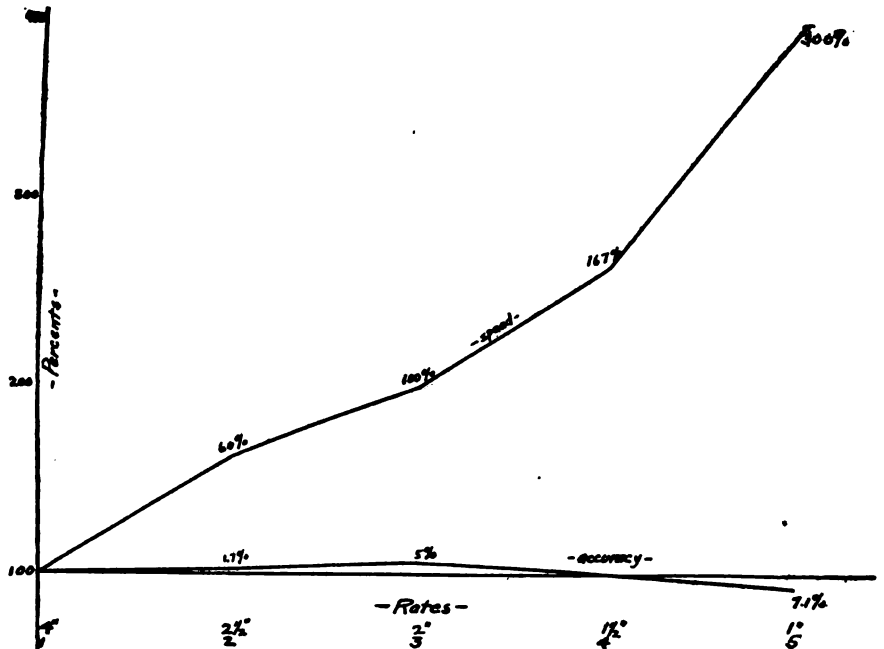
Rate 5. 1" 2100 Lifts.

	88		92		96		100		104		108		112		Tot.		%R
Sub.	R	W	R	W	R	W	+	−	R	W	R	W	R	W	R	W	R
Mz	37	13	35	15	29	21	35	15	28	22	43	7	42	8	214	86	71.3
Ad	29	21	28	22	25	25	36	14	32	18	36	14	40	10	190	110	63.3
Dw	45	5	42	8	34	16	23	27	24	26	39	11	46	4	230	70	76.7
Wn	40	10	36	14	29	21	25	25	37	13	41	9	44	6	227	73	75.7
Ant	33	17	25	25	23	27	38	12	39	11	42	8	45	5	207	93	69.0
Sc	34	16	29	21	20	30	35	15	40	10	41	9	44	6	208	92	69.3
Sum	218	82	195	105	160	140	192	108	200	100	242	58	261	39	1276	524	
%R	72.7		65.0		53.3		64.0		66.7		80.7		87.0		70.9		
AD	8.8		10.3		8.0		10.8		10.7		3.8		3.3		PE(Av.)=1.60		

The PE's of the differences show a pretty definite tendency for the accuracy record to reach its maximum point, in the case of the three weights heavier than 100 gms., at Rate 3, though the difference between this rate and Rate 4 is not very great—6 chances in 7. In the case of three weights lighter than the standard, there is not much evidence of a real change in the accuracy of discrimination from Rate 1 to Rate 4; such evidence as there is, is in favor of Rates 2 and 3. The decrease in accuracy at Rate 5 is, however, clearly present for both the heavier and the lighter weights.

In the diagrams up to the present, no account has been taken of the changes in the lifting rate which parallelled the changes in accuracy; the rates have been merely taken as equal lengths on the baseline. Accordingly, in Diagram 3, the corresponding changes in the two factors have been represented, measurement of both speed and accuracy being made from Rate 1 as 100%. With Rate 1 taken as 100%, Rate 2 becomes 101.7, Rate 3, 104.98, Rate 4, 100.0, and Rate 5, 92.90. It is clear that an increase in the lifting rate of 60%, at Rate 2, produced an increase of 1.7% in accuracy, while an increase of 100% in the rate of lift leads to an increase of 5% in accuracy. At Rate 4, with the speed increased 167%, the accuracy fell back to the level of Rate 1; and at Rate 5 an increase of 300% in lifting rate caused the accuracy to drop 7.1% below the accuracy level of Rate 1. This diagram, then, shows that while speed and accuracy in lifted weights are directly, though not proportionally, related up to a certain point, that when this point is passed the two variables are inversely related, increases in the rate of lift producing decreases in the accuracy record.

Before attempting to summarize the first part of the problem, the question of change in the accuracy of discrimination with change in lifting rate, it is well to consider two facts: (1) that the range of accuracy in the discrimination of weights is limited; and (2) that the difference between the slowest and the fastest rates is absolutely, though not relatively small. It must not be overlooked, also, that Diagram 2 shows a very definite tendency, mentioned before, for the curves of the comparison weights to follow the same course: a gradual rise and fall. In regard to the first statement: we should expect to get 50% of our comparisons right by chance, so that the extreme limits for accurate discrimination lie be-



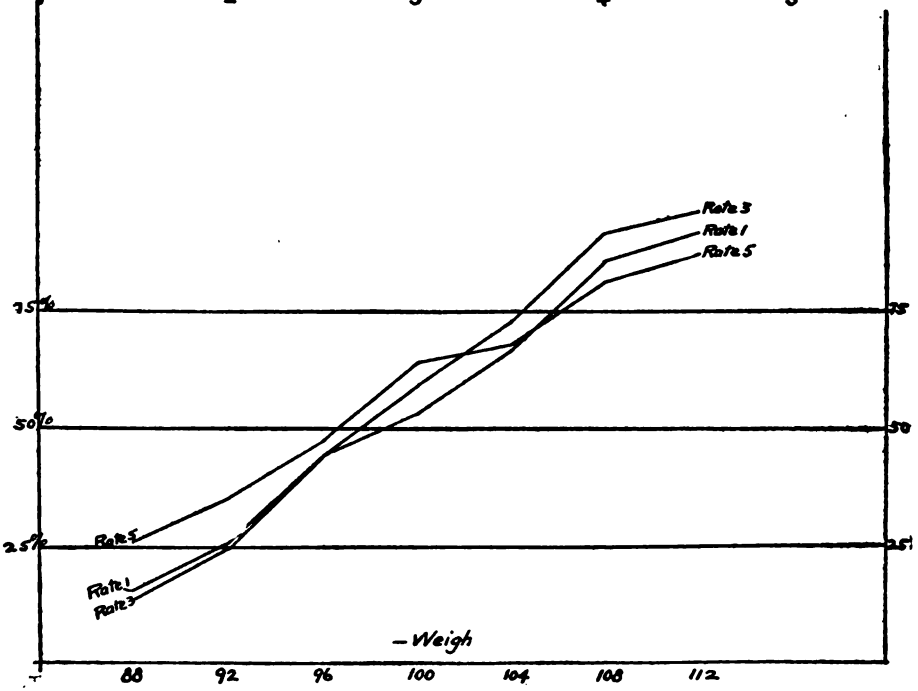


Diagram 3.—Showing the relative changes in accuracy with increased speed. Both speed and accuracy at Rate 1 taken as 100. Accuracy curve is combined records of 6 O's and represents 2100 lifts at each rate.

Diagram 4.—Showing the course of the "heavier" judgments for Rates 1, 3, and 5, for each comparison weight.

tween 50% and 100%. As a matter of fact, all of the weight-pairs, especially the ones containing the lighter weights, have a much smaller range than this at any given rate. Hence, we should not be justified in expecting very large variations in right judgments, say 10% to 90%, no matter what the rate. In regard to the second statement, the range of lifting times was intended to include only those rates which are practicable: a range from say 10 min. interval down to a point below Rate 5, might have shown a greater range in accuracy, but rates much below Rate 1 are far too slow to be usable, and, since the accuracy tends in the other direction, are useless. And on the other hand, it is almost impossible to lift faster than the rate used at Rate 5. To sum up: the differences which appeared between the rates, though small, are about as large as could reasonably be expected. A change in accuracy of 25% from one rate to another would have been indicative of error, and highly suspicious, rather than indicative of the influence of speed of lift. The "definite tendency" for the curves of the comparison weights to follow the same course, offers evidence that the tendency could hardly have been due to chance factors; but that such changes in accuracy as did appear are due to the increased rate at which the weights were lifted.

The answer to the first question is, therefore, as follows. For weights around 100 gms. as standard, it is safe to conclude (1) that there is good evidence of a real change in the accuracy of discrimination as the lifting time decreases from Rate 1 to Rate 5; (2) that the lifting time can be reduced from 4″ per weight with a 4″ interval to 1½″ per weight with a 1½″ interval with little or no loss in accuracy of comparison; (3) that the chances favor Rates 2 and 3 as the "optimum" lifting rates with the balance in favor of Rate 3.

One further aspect of the first question remains to be considered. It is agreed that whether we accept the Müller-Schumann[54] theory of rapidity of lift as the essential factor in judgment, or the Fullerton-Cattell[55] conception of the force of the lift, eg. in muscular and cutaneous sensations from the hand and fingers, as the chief determinant, that accurate discrimination certainly depends on the correct placing, on a

[54]Uber die psych. Grundlagen der Vergleichung gehobenen Gewichte, Archiv. fur d. ges. Physiol. 1889, *45*, 37ff.

[55]On the Perception of Small Differences, 1892, 132.

subjective scale of values, of the impressions got from the first and the second stimuli. How is this evaluation made? And how are we to think of an "optimum rate" in terms of the judgment consciousness?

The judgment process, in those situations in which we are called upon to compare a present sensation with one immediately preceding it, has been a much discussed question. Many of the earlier investigators[56] assumed that the basis of judgment in such cases is the comparison of the visual, muscular, tactual, or other sensations got from the second impression, with the memory image of the preceding one. Bentley,[57] however, in a study of memory image fidelity for brightnesses and colors, reported judgments in which introspection failed to find a trace of comparison of the two stimuli. In these "free judgments," decision came like a flash, and confidence in the correctness of the decision was often quite strong, even though the subject could not state how he made the judgment. Angell[58] reports an experiment with clangs, the object of which was to test the validity of the memory-image theory of comparison. His results showed that for time intervals from 1″ to 60″, there was but little falling off in the accuracy of judgment with increase in time for stimuli differing from 8 to 4 vibrations. Distractions, eg. addition, counting metronome beats, reading backwards, or aloud, applied during the interval did not affect the accuracy of judgment. Angell concludes that for the most part no comparison of standard and variable took place. Whipple[59] in a study of clangs and tones in which the interval between the first and the second stimuli was varied from 2 to 60 sec. also reports immediate judgments in which there was apparently no image present. Pillsbury[60] says "it has been noticed repeatedly that in comparisons, the first standard impression is not ordinarily in mind when the comparison results."

The most careful study of this problem as regards weight lifting has been made by Martin and Müller.[61] These authors

[56]Bentley: The Memory Image and its Qualitative Fidelity, A. J. Psy. 1899, *40*, 8.

[57]Op. cit., p. 39-40.

[58]Discrim. of Clangs for Different Intervals of Time, Amer. J. Psy., 1900, *11*, 58-79.

[59]An Analytic Study of Memory Image, and the Process of Judgment in Discrimination of Clangs and Tones, A. J. Psy., 1901, *12*, 409-457.

[60]Attention, 1908, 178.

[61]Zur Analyse der Unterschiedsempfindlichkeit, 1899, p. 43-44.

state, on the basis of a long research, that what actually takes place in weight comparisons, is that, in a long series the subject forms an "absolute impression" of the stimuli, more often of the variable than the standard. The weights are then judged heavier or lighter in terms of this Absolute Impression, of heaviness or lightness just as we speak of a clear day or a loud noise without comparison with any previous day or noise. Hayden,[62] in later work with lifted weights, accepts this conception of "set" or adjustment in the presence of which the second stimulus is received. Hayden emphasizes the verbal image as the impulse to the adjustment; through such an image, the standard weight is evaluated,—placed on a scale of values.

One further study should be noted. Fernberger,[63] using the Method of Constant Stimuli, has made probably the most careful introspective study of the factors present in judgments of comparing. Fernberger used weights, lines, sounds, and brightnesses. His subjects were required to give a careful description of everything which appeared in consciousness during the process of comparing; to state, as far as possible, on what factors judgment depended and how decision was reached. Fernberger concludes that "the structural components of the process of comparing, we found to be primarily sensations and images." The pattern of the comparing process differed with materials presented, individuals, and the stage of practice reached. Two main stages in the comparison of weights were distinguished. In the early stages, it is essentially the comparison of the imaginal representation of the first stimulus with a perception of the second stimulus. This "imaginal representation" consists of images of different modalities; always, however, there were kinaesthetic images present. In the second stage, a "set" or adjustment is recognized in the presence of which the comparison stimulus is evaluated. It is interesting to see how Fernberger reconciles the Martin and Müller-Fullerton and Cattell controversy on the basis of these two stages of practice. Fullerton and Cattell's subjects, he holds, being relatively untrained, remained largely in the first stage of practice; motor set was absent, therefore, or, if present, only to a slight degree.

[62] Memory for Lifted Weights, A. J. Psy., 1906, *17*, 497-521.

[63] An Introspective Analysis of the Process of Comparing, Psy. Monogr. 1919, *26*, 6.

Judgments were based, then, on kinaesthetic and tactual images of the first stimulus compared with the sensations arising from the second stimulus; that is to say, on the force necessary to hold the weights. Martin and Müller's subjects, being trained, were in the second stage of practice where motor set is the rule; judgments, under these conditions, are determined by the speed and height of lift. Fernberger believes that Fullerton and Cattell were justified in claiming that Müller's contention that judgments are based on the speed and height of lift, are not always true. It is untrue only for the early stages of practice, however.

The introspections of the subjects in the present experiment should be considered in this connection. As none of the subjects were particularly trained in introspection, their reports are rather meager, and are suggestive rather than conclusive. In calling for introspections each subject was asked specifically to state on what criteria his judgments were based, which rate seemed most "comfortable," how confident he felt in his judgments at the different rates, etc. This was done in order to secure definite information; suggestions, as far as possible, were avoided.

INTROSPECTIONS

Mz.—My judgments were based, I think, on muscle and touch sensations from wrist, forearm, and fingers. I had no images. Rate 1 seemed slow compared to the other rates, and I had difficulty in holding my attention to the task. Rates 2 and 3 were better, and I felt surer of myself. I was hurried at Rates 4 and 5, and rather uncomfortable.

Ad.—I found it easier to give a judgment at the slower rates; guessed some at Rate 4 and more at Rate 5. I believe that the slow lifting rate at Rate 1, and the long interval, really served as a distraction rather than as an aid. Often I found my mind "wandering" and my attention on something else. My judgments seemed based on hand and wrist sensations; often they were mechanical.

Dw.—I am not very sure how I compared the weights: muscle and touch sensations from the arm and the hand, I suppose. I preferred the faster rates. Rate 1 was too "dragging," and Rate 5 too strenuous to follow. I felt about the same degree of confidence throughout, except at Rate 5.

Wn.—I think that my comparisons were based on the pres-

sure necessary to hold the weight as I lifted it. Wrist and arm sensations probably came in also. I preferred Rates 2 and 3, and felt more confident there. Rate 1 was too monotonous, and Rate 5 too fast.

Ant.—I can't say exactly how I compared the weights. I think that sensations from wrist and fingers were the basis of judgment. I preferred Rate 2. I felt fairly confident at the first four rates, but very little so at Rate 5.

Sc.—I tried to compare the muscle sensations from the second weight with those from the first. I preferred Rates 2 and 3. Rate 1 was pretty slow, especially after the other rates, and it was hard to keep the attention fixed. I felt more confident at the first three rates.

The bulk of the evidence on the question seems to make it clear that, in the discrimination of lifted weights, a memory image is not usually present, but that, in time, the one impression creates a definite set or adjustment in terms of which the other impression is evaluated. The "optimal rate" is then explicable in terms of duration of memory or set. When the second weight follows the first too quickly, there is a confusion of the factors involved in judgment, i.e. the "psychic factors" of Martin and Müller;[64] when the second weight follows the first too tardily, there is a loss of set, a lapse of attention, and, introspection shows, the introduction of irrelevant factors. Commenting on the influence of the time factor in such judgments, Fullerton and Cattell[65] say:—

"In making experiments on the perception of Small Differences, the time elapsing between the sensations to be compared should not be neglected. In the experiments hitherto described (weights and lights) the interval was always one second. It is possible that two seconds or longer might be a more favorable interval for comparison, but were the interval further lengthened, the first sensation might be expected to fade from memory." Kinnaman[66] also writes: "Judgments of weights are based on focal and marginal sensations, which change greatly as the scale of weights is ascended. It is precisely this change of sensations occurring from the beginning of the lift until the weight clears the support, that we compare and

[64] Op. cit. p. 197.

[65] Op. cit. p. 147.

[66] A Comparison of Judgments for Weights Lifted with the Hand and Foot, A. J. Psy. 1900, *12*, 256-263.

TABLE IV

Discrimination of Lifted Weights when Lifting Time and Interval are varied independently:—2100 Lifts at each Rate:—6 Observers.
(100 gm weight not included in totals)

Rate 6 — 1½" for lifting each weight—3" interval between weight pairs — 2100 Lifts.

	88		92		96		100		104		108		112		Tot.		%
Sub.	R	W	R	W	R	W	+	−	R	W	R	W	R	W	R	W	R
Mr.	44	6	34	16	30	20	28	22	28	22	36	14	46	4	218	82	72.7
Ad	40	10	33	17	33	17	34	16	29	21	39	11	41	9	215	85	71.7
Dw	42	8	38	12	28	22	30	20	34	16	42	8	42	8	226	74	75.3
Wn	42	8	39	11	38	12	30	20	33	17	41	9	44	6	237	63	79.0
Ant	36	14	34	16	30	20	34	16	40	10	45	5	46	4	231	69	77.0
Sc	44	6	36	14	34	16	28	22	30	20	38	12	44	6	226	74	75.3
Sum	248	52	214	86	193	107	184	106	194	106	241	59	263	37	1353	447	
%R		82.7		71.3		64.3		61.3		64.7		80.3		87.7		75.17	
AD		4.4		4.0		5.7		4.3		6.7		5.0		3.0		PE(Av)=1.20	

Rate 7 — 3" for lifting each weight—1½" interval between pairs — 2100 Lifts.

	88		92		96		100		104		108		112		Tot.		%
Sub.	R	W	R	W	R	W	+	−	R	W	R	W	R	W	R	W	R
Mz	42	8	36	14	25	25	32	18	35	15	40	10	44	6	222	78	74.0
Ad	41	9	32	18	28	22	28	22	32	18	40	10	43	7	216	84	72.0
Dw	46	4	42	8	32	18	32	18	38	12	42	8	40	10	240	60	80.0
Wn	34	16	34	16	30	20	32	18	40	10	46	4	46	4	230	70	76.6
Ant	42	8	36	14	28	22	30	20	39	11	47	3	47	3	239	61	79.7
Sc	44	6	35	15	34	16	25	25	38	12	45	5	46	4	242	58	80.7
Sum	249	51	215	85	177	123	179	121	222	78	260	40	266	34	1389	411	
%R		83.0		71.7		59.0		59.7		74.0		86.7		88.7		77.17	
AD		5.3		4.3		4.7		4.7		4.7		5.3		4.0		PE (Av) = 1.32	

judge, rather than the totality of sensation after the weight is up. The reagent will judge best when the speed and other conditions are most favorable for his taking these changes into full consciousness. If the normal speed is maintained, the height of lift is of little consequence provided the observer holds firmly to the changes which occurred at the beginning of the lift. But when the observer defers judgment, he is liable to question his impressions, and do little more than guess."

The second part of the problem has not yet been considered: namely, the influence of the interval between weight pairs on the accuracy of discrimination. This question has been investigated in a second experiment, the results of which are incorporated in Table IV. The results of the previous experiments have indicated a "fall off" in accuracy at Rate 4 for all the comparison weights, and it is important to know whether this loss is due to the increased lifting rate, to the shortening of the interval between weight pairs, or possibly to both factors. In the first experiment of Table IV, 1½" were given for lifting each weight as in R-4, but the interval was increased to 3"—doubled. For simplicity, I have called this Rate 6. In the second experiment of Table IV, the lifting time was increased to 3" for each weight, and the interval decreased to 1½". This combination I have called Rate 7. In the following table, I have compared the reliability of the differences between the accuracy records for Rates 4 and 6, 2 and 7, and 6 and 7.

Table V

	Rate 6	Rate 4	Rate 7	Rate 2
Aver	75.17	76.30	77.17	77.60
PE-Av.	1.20	1.64	1.32	1.78
Diff. Rate 7 & 6			2.00	
Diff. Rate 4 & 6)		1.13		
Diff. Rate 2 & 7)				.43
PE-diff. Rate 7 & 6)			1.78	
PE-diff. Rate 4 & 6)		2.03		
PE-diff. Rate 2 & 7)				2.18

Probability that true diff. between Rates 7 & 6=0 or less=2210 in 10,000 (Roughly 1 in 5)
" " " " " " 4 & 6=0 or less=3528 in 10,000 (Roughly 1 in 3)
" " " " " " 2 & 7=0 or less=4464 in 10,000 (Roughly 1 in 2)

The comparison of Rates 4 and 6 shows the difference between the two rates to be very slight—practically negligible. With a difference in average of 1.13 and a PE (diff.) of 2.03, the chances of a real difference are about 2 in 3 in favor of Rate 4. Evidently the lengthened interval did not produce greater accuracy. When Rate 2 is compared with Rate 7, the chances are even of a true difference; here again the decrease in interval from 2½″ to 1½″ with approximately the same lifting rate, does not affect the accuracy of discrimination. When Rates 7 and 6 are compared, the chances are seen to be about 4 in 5 that Rate 7 is better than Rate 6. As between the two factors, therefore, of time of lift and interval between weight pairs, the evidence is in favor of time of lift as being much more important an influence in accurate discrimination than the interval. In fact, the interval seems to have little or no effect on accuracy.

In all the previous experiments no matter what the rate or the interval the experimenter had noticed that the observers nearly always gave their judgments immediately on lifting the second weight. After the experiment just described, the subjects were asked for an opinion on the influence of the interval. Their introspections bear out the results already given above: all were agreed that, except in a very few cases, their decisions were made as soon as the second weight was lifted, and that the interval had no influence unless very long or very short. In the very long interval it was hard to keep the attention from wandering, while in the very short one, i.e.—one second, there was hardly time to give the verbal response "heavier, a," or "lighter, c," on one weight pair before another pair appeared. Apparently the interval is not a judgment interval at all, but is nothing more than a necessary interim between successive weight pairs.

Several minor questions which arise from the main experiment remain to be considered. One of these is the question of the threshold. Table VI shows the changes in the "general threshold" at the different rates,—the threshold being taken as that difference which the observer can be expected to distinguish 75 times in 100. The entries in the table have been calculated from the Fullerton and Cattell[67] table, and measure directly the accuracy of discrimination. The records of the heavier and the lighter weights have been combined and the

[67]Op. cita p. 16.

Table VI

Threshold (75%) Values for Six Subjects at each of 5 Rates:—2100 Lifts at each rate: 10,500 total.

Rate 1

Obs	96 104	92 108	88 112	Av.	Ad.
Mz.	8.6	6.4	5.6	6.9	1.2
Ad.	9.0	8.0	6.7	7.9	.8
Dw.	11.8	7.0	9.2	9.3	1.6
Wn.	15.1	5.8	7.8	9.6	3.7
Ant.	5.9	4.7	5.7	5.4	.5
Sc.	10.5	5.2	5.2	7.0	2.4
Av.	10.2	6.2	6.7	7.7	1.7
Ad.	2.3	1.0	1.2	1.3	

Rate 2

Obs	96 104	92 108	88 112	Av.	Ad.
Mz.	11.4	5.4	9.2	8.7	2.2
Ad.	10.5	9.5	8.1	9.4	.8
Dw.	5.6	5.2	6.2	5.7	.4
Wn.	11.8	5.0	4.5	7.1	3.1
Ant.	10.6	4.1	4.0	6.2	2.9
Sc.	6.1	3.4	7.2	5.6	1.4
Av.	9.3	5.5	6.5	7.1	1.3
Ad.	2.4	1.4	1.6	1.3	

Rate 3

Obs	96 104	92 108	88 112	Av.	Ad.
Mz.	6.1	4.8	5.0	5.3	.5
Ad.	7.5	8.0	9.2	8.2	.6
Dw.	9.0	5.3	3.7	6.0	2.0
Wn.	3.9	4.6	4.6	4.4	.3
Ant.	9.8	3.9	4.8	6.2	2.4
Sc.	6.9	4.5	5.3	5.6	0.9
Av.	7.2	5.2	5.4	6.0	1.1
Ad.	1.6	1.6	1.2	0.9	

Rate 4

Obs	96 104	92 108	88 112	Av.	Ad.
Mz.	10.5	7.0	7.4	8.3	1.5
Ad.	9.5	12.0	6.9	9.4	2.6
Dw.	5.8	7.0	7.8	6.9	.7
Wn.	8.3	4.3	5.0	5.9	1.6
Ant.	6.0	6.3	4.7	5.7	.6
Sc.	13.3	6.7	5.9	8.6	3.1
Av.	8.9	7.2	6.3	7.5	1.7
Ad.	2.2	1.6	1.1	1.3	

Rate 5

Obs	96 104	92 108	88 112	Av.	Ad.
Mz.	15.4	6.7	10.0	10.0	2.9
Ad.	15.4	14.8	15.1	15.1	.2
Dw.	13.3	6.1	6.0	8.5	3.2
Wn.	6.3	7.3	8.0	7.2	.6
Ant.	9.0	12.3	10.5	10.6	1.1
Sc.	10.6	9.6	10.0	10.1	.4
Av.	11.7	9.5	10.0	10.1	1.4
Ad.	3.0	2.8	2.0	1.8	

Table VII

Constant Errors (Time) for Six Subjects at each of 5 Rates:—2100 Lifts at each rate:—10,500 total.

		Rate 1		
	96	92	88	
Obs.	104	108	112	Av.
Mz.	2.6	–.7	7.2	3.0
Ad.	1.3	3.8	2.0	2.4
Dw.	–.5	.7	–1.5	–.4
Wn.	3.9	3.2	–.5	2.2
Ant.	2.8	4.3	2.8	3.3
Sc.	.0	–.4	.0	–.1
Av.	1.7	1.8	1.7	1.7
		Rate 2		
	96	92	88	
Obs.	104	108	112	Av.
Mz	3.8	6.0	–1.6	2.7
Ad	–2.4	3.0	1.0	.5
Dw	–1.8	–2.0	–2.0	–2.0
Wn	3.1	3.6	3.5	5.1
Ant	5.8	2.4	4.0	4.1
Sc	–.2	–.9	–.5	–.2
Av.	2.2	2.0	0.9	1.7
		Rate 3		
	96	92	88	
Obs.	104	108	112	Av.
Mz	5.1	3.0	5.2	4.4
Ad	–.6	–.4	4.0	1.0
Dw	2.2	2.0	.7	1.6
Wn	2.3	4.0	3.9	3.4
Ant	4.4	5.7	4.4	4.8
Sc	1.4	1.4	1.9	1.6
Av	2.5	2.6	3.4	2.8
		Rate 4		
	96	92	88	
Obs.	104	108	112	Av.
Mz	1.6	3.2	3.5	2.8
Ad	1.2	4.6	4.1	3.3
Dw	.0	1.4	1.5	1.0
Wn	5.0	2.1	3.3	4.3
Ant	4.1	6.6	2.3	4.3
Sc	5.0	1.7	3.4	3.4
Av	2.8	3.3	3.0	3.1
		Rate 5		
	96	92	88	
Obs.	104	108	112	Av.
Mz	–.6	2.7	2.6	1.6
Ad	4.0	4.7	7.5	5.4
Dw	–4.1	–1.0	.5	–1.5
Wn	2.1	1.8	2.0	2.0
Ant	5.8	9.1	6.8	7.2
Sc	8.7	5.2	5.2	6.4
Av	2.7	3.8	4.1	3.5

PE's computed for a given weight difference irrespective of its direction from the standard. There is very little fluctuation in the averages calculated from the weight differences, except for the 4 gm. difference. Here the introduction of the

chance records of the 96 gm. weight is probably the cause, as the Fullerton and Cattell table does not give accurate results for %'s close to 50%. The threshold values for the 5 rates follow the general trend of accuracy, being lowest at Rate 3 and highest at Rate 5. The threshold value at Rate 3, 6, gms., is comparable to that got by Fullerton and Cattell, eg. 6.2 gms., in a series of 4000 lifts with 9 subjects.

The constant errors for each rate are given in Table VII. These values have been calculated according to the method given by Sanford.[69] The averages of the averages shows a tendency, usually found in lifted weight experiments, for the observers to overestimate the second weight. This tendency is evident at all rates and increases fairly regularly as the difference between the weights becomes larger; it is also present in increasing amounts from Rate 1 to Rate 5. With the space error ruled out by the arrangement of the revolving table, we may be inclined to attribute this error to the time factor since the standard always preceded the variable. It is doubtful whether this is more than partly true, however, because in spite of the single time order, S-V, in one-half of the lifts the comparison weight was heavier, and in one-half of the lifts lighter than the standard. Fechner attributed the time error to fatigue, especially with heavy weights, but this explanation is hardly admissible in this experiment. It may very well be that among other factors involved, (particularly with light weights) the tendency of judgment to take a positive form played a part. Hollingworth[70] in studying with 15 different modalities the tendencies and forms of judgment which were preferred when the observer "was left free to select both the direction of judgment (as to first or second stimulus) and the form of expression (positive or negative quality)" reports a "strong tendency to direct the judgment toward the stimulus described as 'positive' in quality" and a "slight tendency to favor the second stimulus presented." It might be, therefore, that as the rate of lifting increased and judgment approximated more closely the "snap" type, the observer if doubtful or "pushed for time" (the direction of judgment being fixed) would tend to say "heavier" rather than "lighter." The increasing number of heavier judgments on the

[69] Op. cita. p. 122.
[70] Experimental Psychology, 1908, p355-356.
[71] Experimental Studies in Judgment, Archiv. Psy., No. 29. p. 67.

100-100 gm. pair, as the rate increased, offers some evidence in favor of this view.

In order to show that the rate of lift is a real factor in determining the accuracy of discrimination, and that the changes in accuracy are not due to the influence of the plus time error alone, I have plotted in Diagram 4 the course of the heavier judgments on each weight for Rates 1, 3, and 5. If the increase in the tendency to overestimate with more rapid lift were the only factor, the curve for Rate 3 should lie above the curve for Rate 1 throughout its course, and Rate 5 above Rate 3: i.e. the plus judgments should increase both for the heavier and the lighter weights. This is not true, however, as the curve for Rate 3 lies below Rate 1 for the lighter weights, and Rate 5 lies below Rate 1 for the heavier weights. The tendency to overestimate alone would have placed Rate 5 above Rates 1 and 3 for the heavier weights and Rate 3 above Rate 1 for the lighter weights. One point further. The tendency to overestimate the second weight and the fact that the point of subjective equality is really nearer the 98 gm. weight than the 100 gm. standard, give the heavier weights an advantage over the lighter ones in that they are actually further from the standard, and hence more easily discriminated. Accordingly, in the following table a comparison is made of the rates, in which the weight differences are taken from 98 gm. as the point of subjective equality, and 112 gm. weight is not

TABLE VIII

Degrees of Confidence for Varying Amounts of Weight Difference at each rate: Six Subjects:—5 Rates:—10,500 Lifts

Rate 1 4″

	88		92		96		104		108		112		Totals	
Obs.	a&b	c	a&b	c	a&b	c	a&b	c	a&b	c	a&b	c	a& b	c
Mz	36	14	39	11	44	6	38	12	37	13	45	5	239	61
Ad.	28	22	30	20	31	19	38	12	38	12	43	7	208	92
Dw.	49	1	45	5	45	5	50	0	47	3	50	0	286	14
Wn.	29	21	23	27	23	27	32	18	30	20	35	15	172	128
Ant.	25	25	19	31	10	40	28	22	43	7	42	8	167	133
Sc.	22	28	21	19	9	41	13	37	23	27	23	27	111	189
Aver.	32	18	29	21	27	23	33	17	36	14	40	10	199	101
Weights	88		92		96		104		108		112			
%R-Cases	84.3		75		56.3		66		85.3		90.7			
%a&b	64		58		54		66		72		80			

Rate 2 2½″

	88		92		96		104		108		112		Totals	
Obs.	a&b	c	a&b	c	a&b	c	a&b	c	a&b	c	a&b	c	a& b	c
Mz.	50	0	49	1	49	1	50	0	49	1	50	0	297	3
Ad.	36	14	31	19	28	22	39	11	35	15	47	3	216	84
Dw.	50	0	50	0	50	0	50	0	50	0	50	0	300	0
Wn.	30	20	23	27	12	38	20	30	36	14	38	12	159	141
Ant.	16	34	21	29	20	30	24	26	47	3	47	3	175	125
Sc.	14	36	9	41	3	47	12	38	35	15	32	18	105	195
Aver.	33	17	30	20	27	23	33	17	42	8	44	6	209	91
Weights	88		92		96		104		108		112			
%R-Cases	87		75.5		58		66.3		89.3		89.7			
%a&b	66		60		54		66		84		88			

TABLE VIII continued

Rate 3 2"

Obs.	88 a&b	c	92 a&b	c	96 a&b	c	104 a&b	c	108 a&b	c	112 a&b	c	Totals a& b	c
Mz.	48	2	50	0	50	0	49	1	50	0	50	0	297	3
Ad.	34	16	31	29	32	18	34	16	43	7	44	6	208	92
Dw.	50	0	50	0	50	0	50	0	50	0	50	0	300	0
Wn.	27	23	28	22	27	23	27	23	36	14	46	4	191	109
Ant.	30	20	18	32	26	24	38	12	47	3	49	1	208	92
Sc.	10	40	16	34	4	46	8	42	27	23	20	30	85	215
Aver.	33	17	32	18	32	18	34	16	42	8	43	7	215	85
Weights	88		92		96		104		108		112			
%R-Cases	86		75.3		56.3		74.7		91		97.3			
%a&b	66		64		64		68		84		86			

Rate4 1½"

Obs.	88 a&b	c	92 a&b	c	96 a&b	c	104 a&b	c	108 a&b	c	112 a&b	c	Totals a& b	c
Mz.	50	0	50	0	50	0	50	0	50	0	50	0	300	0
Ad.	39	11	40	10	46	4	41	9	46	6	48	2	258	42
Dw.	50	0	50	0	50	0	50	0	50	0	50	0	300	0
Wn.	36	14	28	28	26	24	26	24	48	2	45	5	203	97
Ant.	18	32	22	28	16	34	22	28	45	4	47	3	170	130
Sc.	11	39	5	45	3	47	11	39	23	27	24	26	77	223
Aver.	34	16	32	18	32	18	33	17	43	7	44	6	218	82
Weights	88		92		96		104		108		112			
%R-Cases	83.7		68.7		55		70		86		94.3			
%a&b	68		64		64		66		86		88			

Rate 5. 1"

Obs.	88 a&b	c	92 a&b	c	96 a&b	c	104 a&b	c	108 a&b	c	112 a&b	c	Totals a& b	c
Mz.	49	1	49	1	49	1	50	0	50	0	50	0	297	3
Ad.	32	18	38	12	42	8	40	10	35	15	43	7	230	70
Dw.	50	0	50	0	50	0	50	0	50	0	50	0	300	0
Wn.	32	18	20	30	24	26	28	22	37	18	39	11	280	120
Ant.	29	21	26	24	22	28	44	6	44	6	48	2	213	87
Sc.	4	46	1	49	2	48	4	46	25	25	29	21	65	235
Aver.	33	17	30	20	32	18	36	14	40	10	43	7	214	86
Weights	88		92		96		104		108		112			
%R-Cases	72.7		65		53.3		66.7		80.7		87			
%a&b	66		60		64		72		80		86			

	Weights 88-108	Weights 92-104	Weights 96-100
Rate 1	84.8	70.5	54.8
" 2	88.0	71.0	59.3
" 3	88.5	75.0	57.7
" 4	84.8	69.3	59.7
" 5	76.7	65.8	58.7

included. This comparison offers more evidence of a real change in accuracy with increased rate apart from the time error.

There is still an appreciable increase in accuracy from Rate 1, when the effect of the time error is, at least, partly eliminated. On the influence of the time error in general, it might be noted that Table VII shows that the increase in constant error from Rate 3 to Rate 5 is very small—.7; that the lighter weights do not decrease in accuracy from Rate 1 to 3 as they should if increased rate were effective only by increasing the time error; that the further increases at Rates 4 and 5 cause

losses in accuracy which are as large for the heavy weights as for the light weights.

Throughout the course of the experiments, the subjects were required to state their confidence in the accuracy of their comparisons by means of the terms a, b, c. This plan of recording a subject's confidence really combines the Method of Just Noticeable Differences with the Method of R. and W. Cases (vide Fullerton and Cattell, Small Differences, p. 124) on the supposition that a subject's confidence will tend to decrease as the differences between the objective stimuli decrease. Table VIII shows the distribution of a's, b's, and c's for each weight difference at each rate: and a comparison is made of the %R cases for each weight and the % of a's and b's for that rate. Though there is little change in the confidence as the rate increases (in spite of delayed or general introspections to the contrary) the degree of confidence is seen to increase pretty regularly with the % of R-cases, and with the increase in the weight interval. This result is substantial-

TABLE IX

Relation of Correctness of Judgment to Degree of Confidence: Six Subjects=5 Rates–10, 500 Lifts.

Rate 1

	MZ	AD	DW	WN	ANT	SC	AV
a	85	87	86	90	90	100	89
b	72	62	81	75	87	82	76
c	60	68	52	61	63	62	61

Rate 2

	MZ	AD	DW	WN	ANT	SC	AV
a	88	89	81	82	93	—	87
b	56	60	—	88	80	95	76
c	40	59	—	70	68	77	61

Rate 3

	MZ	AD	DW	WN	ANT	SC	AV
a	95	86	82	100	95	—	91
b	76	70	—	83	83	98	82
c	75	63	—	74	59	66	66

Rate 4

	MZ	AD	DW	WN	ANT	SC	AV
a	83	83	77	94	91	—	86
b	50	57	—	78	63	90	68
c	63	67	—	74	63	71	68

Rate 5

	MZ	AD	DW	WN	ANT	SC	AV
a	79	70	76	94	82	—	80
b	53	57	—	83	87	95	65
c	—	60	—	60	57	62	60

Example:—For Mz at Rate 1, 85% of his "a" judgments were correct, 72% of his "b" judgments were correct, and 60% of his "c" judgments were correct.

ly that got by Williamson[71] in a study of the relation between an individual's degrees of confidence and objective differences in the stimuli compared. Williamson had 35 students arrange 25 statements or beliefs by both the Order of Merit Method and the Method of Paired Comparisons. The degree of confidence in the comparisons made by the latter method were required, confidence ranging from 0 to 3 in steps of 1. "0" meant "mere guess" and "3" "absolutely certain." The average confidences for every Rank Difference or "differences discriminated" as found by the Paired Comparisons Method were then correlated with the rank differences by the "squared difference" formula. For the 35 observers the coefficient of correlation ranged from +.705 to +.961 with a median of +.879.

Table IX shows how often the observers were right when *a*, *b*, or *c* was given. They seem to be, on the whole, fairly good judges of the accuracy of their comparisons, as they were more liable to be right when "sure" than when "doubtful." It is evident, too, that an observer is more apt to be right than wrong when he guesses or is doubtful (see Fullerton and Cattell, op. cit. 132).

It is also apparent that those men who were most confident were not always the most accurate:—Mz and Dw who gave practically all "a's." The observers differed much in their understanding of the terms *a*, *b*, and *c*. Dw, for example, was nearly always "sure," while Sc was nearly always "doubtful." Again, Wn and Dw, both of whom gave their judgments quickly and without hesitation were generally better at the faster rates; while Ad, and Ant, who were slow and cautious —inclined to be deliberate—were better at the slower rates, and drop regularly after Rate 3. Williamson[72] also found marked indvidual differences in the range of the correlation coefficients mentioned above, "conservative" persons tending to put a much greater distance between the limits set for confidence degrees, than the more "radical." Martin and Müller[73] classify observers in three groups or types: positive, negative, and indifferent. Those who belong to the "positive" group are strong and muscular individuals, who give a larger % of R cases when the standard is greater than the Variable. Those

[71]Individual Differences in Belief, Measured and Expressed by Degrees of Confidence, Jr. of Phil. Psy. & Scien. Method, 1915, 12, 127-137.

[72]Op. cit. 136-137.

who belong to the "negative" type are relatively weaker and non-muscular and give a larger % of R cases when the standard is lighter than the Variable. (In Martin and Müller's work, the standard came first.) Müller calls this tendency the "type tendency of judgment." In his experiments there were 5 men and 1 woman in the "positive" group, and 5 women and 1 man (of slight muscular development) in the "negative" group. According to this classification, the five observers in this experiment, being all men, should belong to the "positive" group, though the results indicate a larger % of right judgments when the heavier weights came second at each of the rates. This discrepancy in results, may be due to the fact that (1) Martin and Müller's weights were much heavier than the weights used in this experiment, ranging from 416-3,221 gms. and (2) that fatigue and actual muscular effort were absent or negligible among my observers. In general, I believe that the subjects in this experiment may best be classified as (1) impulsive, best at the faster rates, (2) cautious,—best at the slower rates, and (3) indifferent.

CONCLUSIONS

The following conclusions are drawn from this experiment:

1.—There is a real change in the accuracy of discrimination for Lifted Weights as the rate changes from a 4″ lift for each weight, with a 4″ interval between weight pairs, to a 1″ lift for each weight, with a 1″ interval between weight pairs. The Optimal Rate is close to 2″ for each weight with a 2″ interval between weight pairs.

2.—The time-interval (ostensibly given for judgment) is of little or no influence on the accuracy of discrimination, unless it is very long or very short: under both of which conditions it makes for inaccuracy.

3.—The 75% threshold is lowest at Rate 3.

4.—The tendency to overestimate the second weight increases regularly with increase in lifting rate.

5.—Confidence in the accuracy of judgment varies directly with accuracy of discrimination, but does not change much with increase in lifting rate. Different observers are, in general, quick, cautious, or intermediate in giving their decisions.

[7]Op. cit. p. 29ff.

2. Speed and Accuracy in Judging the Comparative Length of Lines.

PROBLEM

In this experiment an attempt was made to find what effect the time given for perception has on the accuracy with which lines of variable length can be judged longer or shorter than a given standard line. In each instance two lines were presented simultaneously, a variable and a standard; the speed, i.e. the interval of exposure, was varied in a regular and predetermined manner, and the accuracy was gauged by the correctness with which the variable line was perceived to be of different length from the standard. The record of the judgments at the different exposure intervals shows whether accuracy for lines tends to increase or diminish with decrease in exposure interval, and indicates, further, which exposure interval is productive of the greatest discriminative clearness. It is of interest to note the relative accuracy with which lines longer or shorter than the standard line are discriminated, and to see, also, how the observer's confidence in his estimates changed with the time allowed for decision. All these questions are considered under the results.

APPARATUS AND PROCEDURE

The exposure material consisted of 12 cards of white cardboard, 14″ by 6″ in size. In the center of each card a horizontal line was drawn, from the left end of which facing the observer a length of 100 mm. was cut off by an upright vertical line 5 mm. long and 2 mm. wide. This 100 mm. length served as the standard line for each card, while the remainder of the line (to the right of the vertical upright) varied in length from 95 mm. to 105 mm. and served as a comparison line to be judged longer or shorter in terms of the standard line. Each card, therefore, contained two lines a variable and a standard. These lines were all of uniform thickness, 1 mm. and, to insure accuracy, were drawn in India ink by an expert draughtsman. The exposure cards, which were rather large, were taken so purposely, in order to prevent judgments being given in terms of the distance of the end of the variable from the edge of the card, rather than the length of the variable itself.

The comparison lines were 95, 97, 98, 98.5, 99, 99.5, 100.5, 101, 101.5, 102, 103, 105 mms. respectively: six longer and six shorter than the standard. These lines were selected because they seem to give a fairly wide range, from differences which can be pretty clearly perceived, to those which do not yield much more than 50% of right determinations.

Different investigators have given different values to the j.n.d. for lines.

E. H. Weber puts the fraction by which one line must differ from another in order to be just perceived, as 1/100 of the standard line for people of keen eyesight. Weber used lines of 50 and 100 mm. as standards. For untrained observers, Weber puts the fraction at 1/25 to 1/20. Fechner reported correct judgments on lines which differ by 1/60 of the length of the standard; and Volkmann differences of 1/90 and 1/100. For a concise survey of these investigations, and of others made by the use of the various psychophysical methods primarily for the purpose of determining the ability of the eye to estimate linear magnitudes, see reference below.[74]

In comparing results got from different experiments with lines, the method of drawing and presenting the lines must be considered, as it makes a good deal of difference whether the lines to be compared are separately drawn, are parallel, or (as in this experiment) are parts of the same line. In the latter case, the judgment is clearly influenced by the presence of the whole line.[75]

For an exposure apparatus, I used a box 14½″ wide, 10″ long, and 6″ high. This box was open at the top and one side, and was large enough for the exposure cards to fit into it easily. The observer was seated about 10 feet before the open end of the box, and on a level with it, his back to a window. Care was taken to secure uniform lighting and to avoid shadows being cast across the face of the cards. Each card was numbered, and a tab fastened to the top so that it could be turned easily by the experimenter. The cards were exposed for a definite time interval, during which interval the observer designated the comparison line as longer or shorter than the standard; and expressed his confidence in

[74] Henmon: Time of Perception as a Measure of Difference in Sensation, Archiv. of Phil. Psy. & Scient. Methods, 1906, p. 54-61.

[75] Münsterburg: Beiträge zur experimentellen Psychologie, 1899, Heft 2,130.

his perception by the letters "a" sure, "b" fairly sure, and "c" uncertain or guess. The method is thus seen to be the same as that used in the previous experiments on lifted weights; Right and Wrong cases, plus the additional requirement that the observer give his feeling of confidence in his answer.

The twelve cards were presented in 20 successive series, the order of presentation being changed each time: thus at a given exposure rate an observer made 20 judgments on each line, or a total of 240 judgments on the 12 lines.

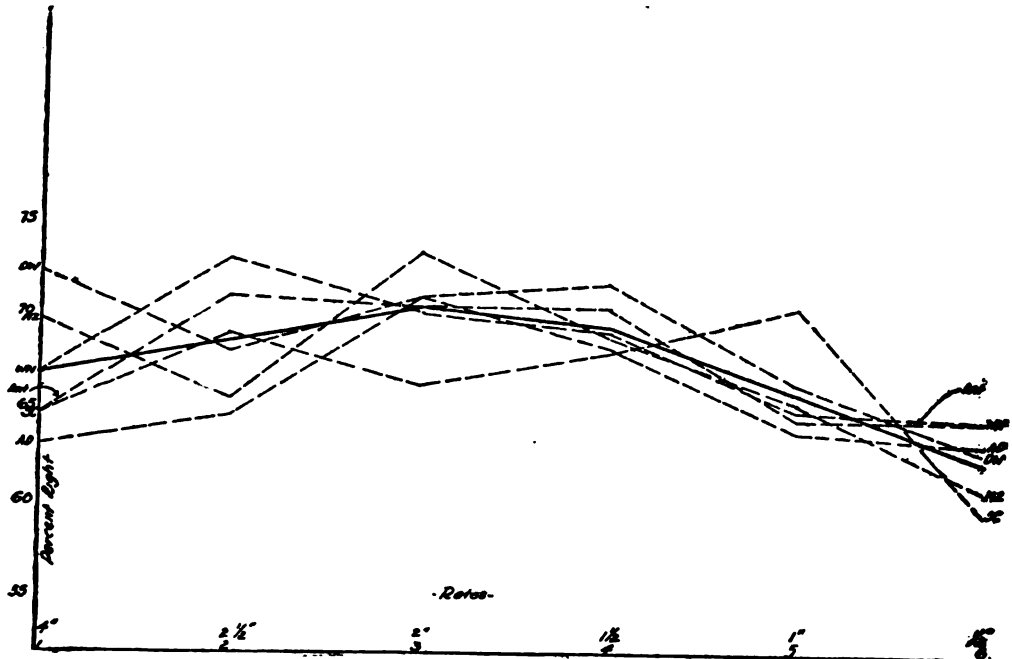
The exposure intervals were regulated by a bell metronome. Six intervals were used, 4", 2½", 2", 1½", 1" ½": the intervals, except the ½", being marked by four beats of the metronome, with bell on the 4th beat. For the last interval, the bell rang on every other beat with the metronome set for 240 (runner removed). On each bell the experimenter dropped a card, thus making the exposure for the determined interval as described above. This procedure was continued until all 12 cards had been shown. Judgments, as they were spoken by the observer, were recorded on a specially prepared blank by a third person. The writer, who acted as experimenter throughout, practiced for a considerable time until he could turn the cards with regularity and smoothness, and there was little or no confusion in the presentation. None of the observers complained of movements sufficient to disturb their perception of the lines as they were successively exposed.

PRACTICE AND FATIGUE

To avoid the massing of whatever practice effect there might be at any one rate of exposure, the series were taken according to a regular plan: 20 at one rate, 20 at another rate, etc. As never more than 180 judgments were recorded at any one sitting, it was not believed that the factors of fatigue or eyestrain need be considered.

RESULTS

The six men, Mz, Dw, Wn, Ad, Sc, Ant, who acted as subjects in the weight-lifting experiments served as observers in the present experiment also. Table X contains the records of each individual observer in % of right cases of all 12 lines; also the % of right cases for each comparison line as well as the total % accuracy for the given rate. The results at any given rate are based on 1440 observations.



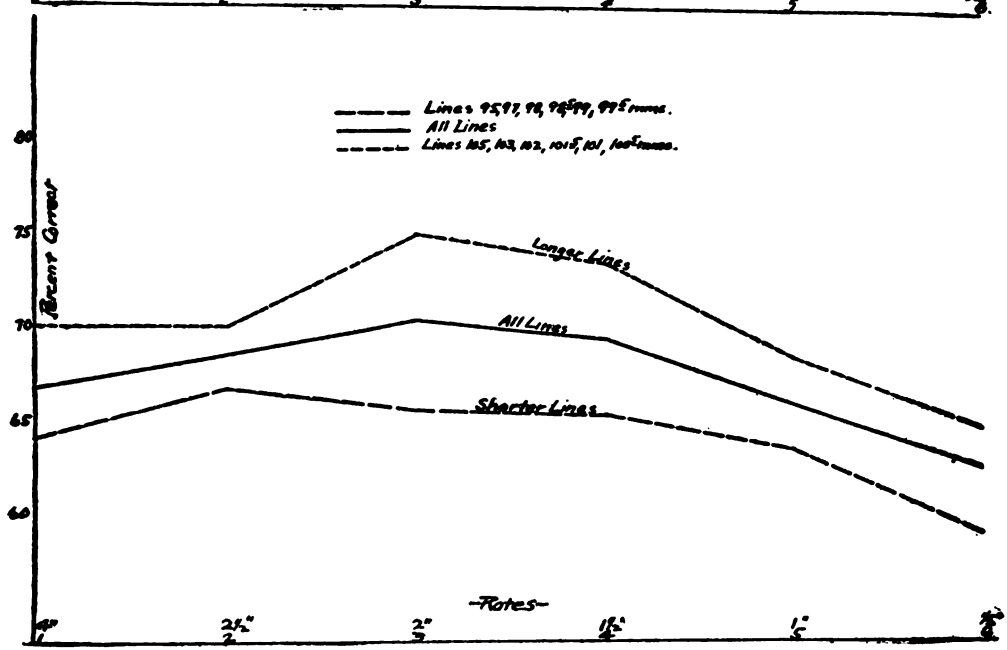


Diagram 5.—Combined accuracy record for 6 O's _____ Also individual records .....1440 judgments at each rate

Diagram 6.—Combined accuracy record for all 12 lines; also for 6 lines longer and 6 lines shorter than the standard 100 mm. line.

Table X

Perception of Lines:—Method of Right and Wrong Cases:—Six subjects, 8640 Observations—1440 at each of six exposure rates.

Rate 1 4″ — 1440 Observations

Obs.	95 R	95 W	97 R	97 W	98 R	98 W	98.5 R	98.5 W	99 R	99 W	99.5 R	99.5 W	100.5 R	100.5 W	101 R	101 W	101.5 R	101.5 W	102 R	102 W	103 R	103 W	105 R	105 W	Tot. R	Tot. W	% R
Mz	17	3	16	4	16	4	10	10	10	10	12	8	10	10	12	8	16	4	16	4	15	5	17	3	167	73	69.6
Dw.	16	4	16	4	17	3	14	6	13	7	15	5	12	8	10	10	12	8	13	7	17	3	18	2	173	67	72.1
Wn.	15	5	12	8	12	8	15	5	12	8	10	10	10	10	14	6	13	7	14	6	17	3	16	4	160	80	66.7
Ad.	12	8	14	6	13	7	10	10	8	12	11	9	13	7	14	6	11	9	14	6	16	4	15	5	151	89	62.9
Sc.	17	3	12	8	12	9	12	8	8	12	9	11	10	10	12	8	16	4	16	4	15	5	17	3	155	85	64.6
Ant.	16	4	14	6	17	3	11	9	10	10	8	12	11	9	14	6	14	6	12	8	14	6	14	6	155	85	64.6
Sum.	93	27	84	36	86	34	72	48	61	59	65	55	66	54	76	44	82	38	85	35	94	26	97	23	961	479	
%R.	77.5		70.0		71.7		60.0		50.8		54.2		55.0		63.3		68.3		70.8		78.3		80.0		66.73		
Ad.	6.7		6.7		11.7		8.3		7.8		9.2		5.0		6.7		8.3		6.2		5.0		5.8				
PE (Aver.)																									1.03		

Rate 2 2½″ — 1440 Observations

Obs.	95 R	95 W	97 R	97 W	98 R	98 W	98.5 R	98.5 W	99 R	99 W	99.5 R	99.5 W	100.5 R	100.5 W	101 R	101 W	101.5 R	101.5 W	102 R	102 W	103 R	103 W	105 R	105 W	Tot. R	Tot. W	% R
Mz.	16	4	15	5	13	7	11	9	10	10	10	10	12	8	12	8	13	7	14	6	15	5	16	4	157	83	65.4
Dw.	18	2	16	4	16	4	12	8	13	7	10	10	11	9	11	9	10	10	12	8	17	3	17	3	163	77	67.9
Wn.	17	3	13	7	14	6	13	7	14	6	11	9	12	8	14	6	16	4	15	5	18	2	18	2	175	65	72.9
Ad.	13	7	12	8	17	3	15	5	10	10	10	10	11	9	10	10	14	6	11	9	17	3	15	5	155	85	64.6
Sc.	16	4	14	6	12	8	15	5	12	8	9	11	10	10	15	5	12	8	16	4	15	5	19	1	165	75	68.8
Ant.	15	5	15	5	14	6	14	6	13	7	12	8	12	8	12	8	15	5	15	5	16	4	17	3	170	70	70.8
Sum.	95	25	85	35	86	34	80	40	72	48	62	58	68	52	74	46	80	40	83	37	98	22	102	19	985	455	
%R.	79.2		70.8		71.7		66.7		60.0		51.7		56.7		61.7		66.7		69.2		81.7		85.0		68.40		
Ad.	6.2		5.8		7.3		6.7		5.8		4.0		3.3		7.3		8.3		7.8		5.0		4.2				
PE. (Aver.)																									.967		

Rate 3 2″ — 1440 Observations

Obs.	95 R	95 W	97 R	97 W	98 R	98 W	98.5 R	98.5 W	99 R	99 W	99.5 R	99.5 W	100.5 R	100.5 W	101 R	101 W	101.5 R	101.5 W	102 R	102 W	103 R	103 W	105 R	105 W	Tot. R	Tot. W	% R
Mz.	18	2	16	4	14	6	12	8	10	10	14	6	14	6	14	6	12	8	15	5	17	3	20	0	176	64	73.3
Dw.	14	6	18	2	13	7	10	10	15	5	15	5	8	12	14	6	14	6	14	6	17	3	19	2	170	70	70.8
Wn.	18	2	16	4	12	8	14	6	15	5	11	9	8	12	12	8	12	8	15	5	19	1	16	4	168	72	70.0
Ad.	15	5	17	3	10	10	12	8	8	12	12	8	14	6	16	4	16	4	14	6	16	4	20	0	170	70	70.8
Sc.	17	3	10	10	14	6	10	10	6	14	10	10	10	10	14	6	16	4	17	3	17	3	18	2	159	81	66.2
Ant.	16	4	15	5	15	5	14	6	10	10	6	14	12	8	12	8	16	4	17	3	18	2	18	2	169	71	70.4
Sum.	98	22	92	28	78	42	72	48	64	46	68	52	66	54	82	38	86	34	92	28	104	16	110	10	1012	428	
%R.	81.7		76.7		65.0		60.0		53.3		56.6		55.0		68.3		71.7		76.7		86.7		91.7		70.28		
Ad.	6.7		9.4		6.7		6.7		14.4		11.7		11.7		5.6		8.3		5.7		4.0		5.7				
PE. (Aver.)																									1.22		

Rate 4 1½" — 1440 Observations

Obs.	95 R	95 W	97 R	97 W	98 R	98 W	98.5 R	98.5 W	99 R	99 W	99.5 R	99.5 W	100.5 R	100.5 W	101 R	101 W	101.5 R	101.5 W	102 R	102 W	103 R	103 W	105 R	105 W	Tot. R	Tot. W	% R
Mz.	14	6	14	6	12	8	15	5	14	6	8	12	10	10	14	6	15	5	15	5	17	3	17	3	165	75	68.8
Dw.	17	3	15	5	13	7	14	6	13	7	10	10	11	9	16	4	15	5	14	6	16	4	18	2	172	68	71.7
Wn.	16	4	14	6	13	7	15	5	10	10	8	12	9	11	15	5	16	4	14	6	17	3	19	1	166	74	69.2
Ad.	15	5	17	3	15	5	13	7	8	12	11	9	10	10	12	8	15	5	13	7	17	3	18	2	164	76	68.3
Sc.	16	4	12	8	14	6	12	8	11	9	12	8	8	12	15	5	15	5	14	6	15	5	19	1	163	77	67.9
Ant.	16	4	15	5	15	5	13	7	12	8	9	11	7	13	14	6	14	6	16	4	18	2	20	0	169	71	70.4
Sum.	94	26	87	33	82	38	82	38	68	52	58	62	55	65	86	34	90	30	86	34	100	20	111	9	999	441	
%R.	78.3		72.5		68.3		68.3		56.7		48.3		45.8		71.7		75.0		71.7		83.3		92.5		69.4		
Ad.	4.0		5.8		5.0		5.0		8.3		6.7		5.8		5.0		1.7		4.0		4.0		5.0				
PE. (Aver.)																									1.14		

Rate 5 1" — 1440 Observations

Obs.	95 R	95 W	97 R	97 W	98 R	98 W	98.5 R	98.5 W	99 R	99 W	99.5 R	99.5 W	100.5 R	100.5 W	101 R	101 W	101.5 R	101.5 W	102 R	102 W	103 R	103 W	105 R	105 W	Tot. R	Tot. W	% R
Mz.	16	4	15	5	13	7	14	6	11	9	9	11	8	12	10	10	12	8	14	6	17	3	18	2	157	83	65.4
Dw.	17	3	13	7	12	8	13	7	10	10	10	10	12	8	12	8	12	8	15	5	16	4	17	3	159	81	66.3
Wn.	18	2	14	6	12	8	10	10	9	11	12	8	6	14	12	8	14	6	15	5	17	3	17	3	156	84	65.0
Ad.	16	4	13	7	14	6	13	7	10	10	8	12	10	10	10	10	14	6	13	7	16	4	16	4	153	87	63.0
Sc.	18	2	14	6	15	5	15	5	12	8	11	9	11	9	14	6	15	5	15	5	14	6	15	5	169	71	70.4
Ant.	16	4	13	7	10	10	12	8	10	10	9	11	9	11	12	8	13	7	16	4	17	3	18	2	155	85	64.6
Sum.	101	19	82	38	76	44	77	43	62	58	59	61	56	64	70	50	80	40	88	32	97	23	101	19	949	91	
%R.	84.2		68.3		63.3		64.2		51.7		49.2		46.7		58.3		66.7		73.3		80.8		84.2		65.9		
Ad.	4.2		3.3		6.6		6.1		3.9		5.8		8.3		5.7		5.0		4.0		4.2		4.5				
PE. (Aver.)																									1.11		

Rate 6 ½" — 1440 Observations

Obs.	95 R	95 W	97 R	97 W	98 R	98 W	98.5 R	98.5 W	99 R	99 W	99.5 R	99.5 W	100.5 R	100.5 W	101 R	101 W	101.5 R	101.5 W	102 R	102 W	103 R	103 W	105 R	105 W	Tot. R	Tot. W	% R
Mz.	16	4	14	6	13	7	14	6	6	14	6	14	8	12	10	10	12	8	16	4	14	6	17	3	146	94	60.9
Dw.	17	3	16	4	14	6	10	10	10	10	12	8	6	14	10	10	14	6	10	10	15	5	17	3	151	89	62.9
Wn.	12	8	10	10	12	8	12	8	12	8	10	10	14	6	14	6	14	6	14	6	13	7	16	4	155	85	64.6
Ad.	13	7	16	4	10	10	12	8	8	12	14	6	10	10	16	4	14	4	12	8	16	4	11	9	152	88	63.3
Sc.	12	8	13	7	10	10	10	10	8	12	6	14	12	8	12	8	14	6	16	4	15	5	15	5	143	97	59.6
Ant.	16	4	12	8	15	5	13	7	10	10	12	8	8	12	12	8	15	5	14	6	12	8	16	4	155	85	64.6
Sum.	86	34	81	39	74	46	71	49	54	66	60	60	58	62	74	46	85	35	82	38	85	35	92	28	902	538	
%R.	71.7		67.5		61.7		59.3		45.0		50.0		48.3		61.7		70.8		68.3		70.8		76.7		62.64		
Ad.	10.0		9.2		8.3		5.8		8.3		13.3		11.7		9.0		4.5		9.0		5.8		7.7				
PE. (Aver.)																									1.13		

At first blush it might seem that the more time an observer is allowed for discriminating the difference between two lines, the more accurate he will be. This is not true, however, at least for the conditions of this experiment, as Table X and Diagram 5 make clear. The course of accuracy rises slightly from Rate 1 to Rate 3; there is little change from Rate 3 to Rate 4, and a fairly definite "slump" in accuracy through Rates 5 and 6. The individual records (Diagram 5) though they show some irregularity, as might be expected, are very close together, the greatest divergences coming at Rates 1 and 2. Reference to Table X shows that for the lines less than 100 mm. only one, 98 mm. was perceived most accurately at Rate 1, this Rate being as good but no better than Rate 2. For lines greater than 100 mm. Rate 1 is nowhere productive of the greatest accuracy.

As brought out in the previous experiment, there are two ways of studying changes in accuracy corresponding to changes in speed, i.e. time given for discriminating between two stimuli. One is to compare the number of accurate responses from the subjects at the different rates, and the other to compare the right cases for the variable stimuli at the different rates of exposure. The second method gives the more detailed information. Both of these methods were used in the weight-lifting experiments. In the present experiment, however, I have combined the six lines shorter than the standard, and the six lines longer than the standard, rather than taking each comparison line separately. This was done because the lines differ so slightly from each other that I did not think it worth while considering each separately. The accuracy curves for the longer and shorter lines as well as the curve for all 12 lines are given in Diagram 6.

Tables XI and XII give the data from which the curves in Diagram II were plotted, and the reliabilities of the differences between the rates.

The PE (true-obt. Av.) were calculated from the formula PE (t-o Av.) $= \frac{\text{PE (dis.)}}{\sqrt{N}}$ and the PE (diffs.) from the formula: PE (tr-obt. diff. A-B) $= \sqrt{\text{PE}^2\text{ (t-oA)} + \text{PE}^2\text{ (t-oB)}}$. Diagram 6 shows that the curve for the shorter lines reaches its highest point at Rate 2, and hence the records for the other 5 rates are compared with it. There is little evidence of any

TABLE XI

(All 12 Lines)

	Rate 1	Rate 2	Rate 3	Rate 4	Rate 5	Rate 6
Aver.	66.78	68.40	70.28	69.38	65.90	62.64
PE (Av)	1.03	.967	1.22	1.14	1.11	1.13
Diff (Av) from R–3	3.55	1.88	——	.90	4.38	7.64
PE (diff) from R–3	1.59	1.56	——	1.67	1.65	1.66

Probability that true diff. between Rates 3 & 1=0 or less=664 in 10,000 (Roughly 1 in 15)
" " " " " " 3 & 2=0 or less=2090 in 10,000 (Roughly 1 in 5)
" " " " " " 3 & 4=0 or less=3553 in 10,000 (Roughly 1 in 3)
" " " " " " 3 & 5=0 or less= 369 in 10,000 (Roughly 1 in 28)
" " " " " " 3 & 6=0 or less= 10 in 10,000 (Roughly 1 in 1000)

TABLE XII

(Lines 95-97-98-98.5-99-99.5)

	Rate 1	Rate 2	Rate 3	Rate 4	Rate 5	Rate 6
Aver.	64.03	66.70	65.60	65.40	63.50	59.20
PE (Av)	1.56	1.34	1.78	1.37	1.47	1.59
Diff (Av) from R–2)	2.67	——	1.10	1.3	3.2	7.5
PE (diff) from R–2)	2.05	——	2.23	1.92	1.99	2.08

Probability that true diff. bet. Rates 2 and 1=0 or less=1003 in 10,000 (Roughly 1 in 10)
" " " " " " 2 and 3=0 or less=3679 in 10,000 (Roughly 1 in 3)
" " " " " " 2 and 4=0 or less=3255 in 10,000 (Roughly 1 in 3)
" " " " " " 2 and 5=0 or less=1388 in 10,000 (Roughly 1 in 7)
" " " " " " 2 and 6=0 or less= 75 in 10,000 (Roughly 1 in 134)

(Lines 105-103-102-101-100.5)

	Rate 1	Rate 2	Rate 3	Rate 4	Rate 5	Rate 6
Aver.	70.00	70.00	75.00	73.3	68.3	64.7
PE (Av)	1.31	1.41	1.61	1.70	1.66	1.58
Diff (Av) from R–3)	5.0	5.0	——	1.7	6.7	10.3
PE (diff) from R–3)	2.07	2.14	——	2.34	2.31	2.25

Probability that true diff. bet. Rates 3 and 1=0 or less=520 in 10,000 (Roughly 1 in 18)
" " " " " " 3 and 2=0 or less=572 in 10,000 (Roughly 1 in 16)
" " " " " " 3 and 4=0 or less=3112 in 10,000 (Roughly 1 in 3)
" " " " " " 3 and 5=0 or less=252 in 10,000 (Roughly 1 in 40)
" " " " " " 3 and 6=0 or less= 12 in 10,000 (Roughly 1 in 800)

precise optimal rate for accuracy, the change from Rate 2 to Rate 4 being so slight as to be almost negligible. The rise in accuracy from Rate 1 to 2, and the "drop" from Rate 4 to

6, are, however, fairly reliable—particularly the last mentioned change.

For the lines greater than the standard, Rate 3 is the point of greatest accuracy, though the chances are small of a difference between this rate and Rate 4. The evidence in favor of Rate 3 is considerably better in comparison with the other rates, as the reliability table shows: the decrease through Rates 5 and 6 being much more definite and probable.

Apparently there is no rate which can be designated, finally, as the "optimal" exposure interval for all 12 lines; though the accuracy is highest at Rates 3 and 4. About all that can be fairly said is that the differences between the variable and the standard lines are perceived as well when 1″ is given for comparison as when 4″ are given; and that when 2½″, 2″, 1½″, are allowed, the comparisons are slightly more accurate than when 4″, 1″, or ½″ are given. If we remember (as pointed out in the experiment on lifted weights) that the range of accuracy, which we may expect, is limited and that a rather large number of judgments were taken at each rate, the balance in favor of Rates 3 and 4 is believed to be relatively high. It is very probable that the fluctuations in accuracy from Rate 2 to Rate 4 are so slight as to show no real difference in the subject's ability to discriminate at these different intervals: in fact, they suggest chance factors.

Diagram 7 shows the corresponding changes in speed (decreased interval and accuracy (perception of differences). With Rate 1 as unit, 100%, Rate 2, measured from Rate 1, becomes 102.6, Rate 3, 105.3, Rate 4, 103.9, Rate 5, .987, and Rate 6, .938. It is evident that an increase in speed of exposure of 60% produced an increase of 2.6% in the accuracy of perception; an increase of 100% in exposure rate produced a corresponding increase of 5.3% in accuracy measured from Rate 1; an increase of 16% in speed produced an increase of 3.9% in accuracy. At Rate 5, however, an increase of 300% in exposure rate caused the accuracy to fall 1.3% below the level of Rate 1, while at Rate 6, with the speed increased 700%, the accuracy decreased 6.2% from the level of Rate 1. As in the case of lifted weights, speed and accuracy are directly though not proportionally related, up to Rate 3,—an interval of 2″; after this point the two variables are inversely related, a decrease in exposure interval—i.e. increase in rate—causing a decrease in accuracy.

The results outlined above are very similar to those got in the weight-lifting experiment, and the explanation must be very much the same in both cases. Probably what actually happens is that the observer's decision is made as soon as his eyes "run over" the lines, and that the time left is simply unused. When much time is left, fluctuations of attention, monotony, irrelevant details, etc., creep in; when the cards follow each other very rapidly, confusion ensues, due partly to the fact that one vocalization follows another so quickly that the shift from one to another is difficult—and oftentimes not made. Other factors which may be responsible for changes in accuracy have been described by Müller[76] as "casual" or "incidental" errors. Some of these, called "psychological errors," apply well here. Such errors are due to (1) varying degree and direction of attention; (2) variation among the elements of the complex determining judgment; (3) degree and direction of expectation; (4) influence of the order of the preceding experiments. Errors in the last named are due to slip comparisons, shift of standard, change in the concentration of attention, and inertia or habituation of judgment.

The "optimal" exposure interval is, therefore, that interval which permts the observer to exploit the lines, and speak his judgment without undue haste, or too long a lapse before the next card appears. The introspections of the observers, as well as the observations of the experimenter confirm this statement. Frequently, at Rate, I noticed a tendency on the part of the subject to hurry the experiment along—judgments were spoken as soon as the card was exposed, and often the observer's eyes wandered about the room coming back with a jerk when the bell rang for the appearance of another card. On the other hand, at the conclusion of a series of 20 comparisons at Rate 6, an observer would often say: "I meant to say 'shorter, a' in one of those last judgments, but I simply said 'longer, a' before I could stop myself." With a number of possible responses (six in all) all in equal readiness, it is easy to understand how these "false reactions" could happen.

The introspections of the subjects in this experiment are given subject to the same qualifications made for the introspections in the previous experiment.

[76]Müller: Die Gesichtpunkte und die Tatsachen der psychophysichen Methodik, 1904, 109-112.

INTROSPECTIONS

Mz.—Had no general method. Moved my eyes across the cards, usually from left to right, and decided in term of the first impression. About the same degree of confidence throughout.

Dw.—I used no particular method at the slow rates. At the fast rates, tried to get an impression of the standard line and compare in terms of it. Occasionally I glanced back at the standard when uncertain; not much change in confidence.

Wn.—Varied by method considerably; finally settled on fixating the central dividing line, and glancing quickly to the left and right. Usually seemed to have a pretty good idea of the standard line, and did not look at it often. I was less confident at the shorter intervals.

Ad.—Used quick eye movements. I had no particular plan so far as I know. Was less confident as the intervals grew shorter.

Sc.—Usually took a quick glance from left to right. Guessed sometimes at the small differences. Never felt particularly confident; used "b" when I was pretty sure.

Ant.—I soon seemed to have a pretty clear idea of the standard. Then I glanced quickly at the comparison line, and" checked" up by looking back at the standard if uncertain. Was fairly confident, though I guessed some times at the short intervals.

In experiments on Lifted Weights the stimuli are presented successively and the second stimulus is always compared with the one immediately preceding. In this experiment the task set the observer is different, in that the standard and the variable are always presented simultaneously and comparison is made directly. No "set" or "absolute impression" is necessary, though the observer often forms a more or less clear visual impression of the standard (see Fernberger[77]). Such an "impression" can hardly be an influential factor, however, in an experiment like the present, as in experiments on lifted weights: for the observer can always "refresh" his memory, when uncertain, by a quick glance at the standard. Accuracy, therefore, must depend largely (after allowance is made for the personal equation) on visual acuity, attention to the task,

[77]An Introspective Analysis of the Process of Comparing, Psy. Monogr., 1919, 103-110.

practice. Münsterburg[78] finds that in the estimation of simultaneously presented lengths and average error varies from 1.1% to 2.3%, when the eyes are permitted to move freely, while the average error varies from 3.7% to 4.9% when the eyes remain fixed. He takes this as clear proof of the dependence of our estimates of linear magnitudes on muscular sensations.

One source of error in this experiment, which has been touched upon, but not fully considered, lies in the fact that the observers were required to give their answers aloud, rather than express them by releasing a key, raising a hand or by some other mechanical device. This was certainly a source of error at Rate 6, where shift from one vocal set to another was difficult Still it could not have been the only source of error, here or elsewhere, as the accuracy increased appreciably from Rate 1 to Rate 3: from an interval of 4″ to one of 2″. None of the subjects had any apparent difficulty in speaking their decisions at Rates 4 or 5, and as a matter of fact Rate 5 was as accurate at Rate 1. It is fully recognized that the vocal recording of judgments restricts the exposure interval, as it would be impossible to speak a judgment when the cards are exposed, say for ¼″. However, the method used in this experiment precluded the use of exposure rates below ½″, as it was mechanically impossible to present the cards at a rate faster than that, while it was not the object of the experiment to find the shortest interval at which judgment could be made, but only to trace the accuracy of spoken judgments over a fairly wide range of exposure intervals. The decrease in accuracy from Rate 3 on made the use of rates faster than 6 give little promise of an upward trend of the accuracy curve; while the rise in the accuracy curve from Rate 1 did not give much hope of greater accuracy below the slowest rate used. It is worthy of note, too, that since all six observers used only "c" at Rate 6, that the vocal response was not much more difficult than releasing a key.

Diagram 6 shows that the lines greater than 100 mm. were perceived more accurately throughout, than those less than the standard. The exceptions, in every case but two, occurred with those lines which differed by .5 mm. from the standard, where the accuracy was always close to 50%. This more accurate estimation of the longer lines may be due to the ten-

[78]Op. cita. p. 178-179.

dency to overestimate the variable as compared with the standard, which Munsterburg reports, and which he explains as due to the observer's tendency to glance quickly and indirectly at the standard line while the variable is examined more closely. These cursory glimpses of the standard, taken oftentimes as "checkups" on one's impression, result in its underestimation.

In this experiment the direction of judgment was fixed; variable always on the right, standard on the left. This, doubtless, introduced a constant space error. But since this error was present at all six rates, it does not vitiate a comparison of results,—which is the real object of the experiment. For this reason a series of judgments with the variable and standard in reverse positions was not taken.

Henmon[79] finds evidence of the more accurate perception of the shorter of two given lines. Using two lines only, 20 mm. and 20.3 mm. he reports a "curious constant error" in that his three observers give quicker and more confident judgments when the direction of judgment was from the longer to the shorter line. The observers reacted one-half of the time to the longer line and one-half of the time to the shorter line. These results are contrary to those got in this experiment, but the two experiments are hardly comparable. Henmon used only two lines, both comparatively short, direction of judgment was not fixed, and the exposure time was variable, exposure continuing until the observer made a judgment. Also, the two lines were separate and not joined as in this experiment. Judgment was expressed by releasing a key, and time of discrimination was recorded by the Hipp Chronoscope.

Table XIII shows the number of times the various degrees of confidence were used by the different observers, and the % of *a*'s and *b*'s compared with the % of right cases for each line difference. As in the weight-lifting experiments, the confidence decreases regularly with the decrease in right cases, but the degree of confidence is hardly an accurate measure of the reliability of an individual's records for accuracy. In spite of instructions as to the meaning of the terms *a*, *b*, *c*, the different observers gave them very different values. Table XIV, for example, shows that Mz was entirely confident at Rates 4 and 5, while his accuracy records at the same rates was 68.8% and 65.8%. Ant and Ad used a majority of *b*'s and

[79]Time and Accuracy of Judgment, Psy. Rev., 1811, *18*, 189ff.

c's while Sc never used *a*, and Wn used *a* only 20 times in 1440 judgments. Henmon[60] reports that one of his observers used *a* 90.4% of the time, while another used *a* only 9.3% of the time. At Rate 6 all of the observers used *c*. This may have been due to a feeling that "I should be less confident" rather than to a real loss of confidence; for accuracy was still fairly high above 50%.

These results indicate the difficulty of getting a real scale for confidence in such measures as these. The relation of confidence to accuracy seems to be an individual affair, and it is impossible to say whether observers who are usually confident are generally more or less accurate than those who are cautious.

Individual differences in the accuracy of comparison were not very pronounced except possibly at Rates 1 and 2. Dw and Wn, both classified in the previous experiment as quick in giving their judgments, and most accurate at the faster rates, are no better than Ad or Ant, both classed as "cautious" in the weight experiment and best at the slow rates. With the possible exception of Sc, the observers seem to possess nearly the same ability to perceive linear differences. In the use of the confidence categories, the observers follow very closely their records in the weight lifting experiment: Mz and Dw use *a* most of the time, Wn and Sc *c*, and Ad and Ant use all three terms with a strong "leaning" to the *c* category.

CONCLUSIONS

1.—Under the conditions of this experiment, the accuracy of perception of differences between the 100 mm. standard and variables ranging from 5 mm. above to 5 mm. below the standard does not change appreciably whether the exposure interval is 2½", 2", or 1½". When the exposure interval is increased to 4" or decreased to 1" there is a relatively slight, but definite, "drop" in accuracy. At ½" exposure there is a pronounced decrease in the accuracy of perception.

2.—The variable lines longer than the standard give higher %s of right cases than those shorter than the standard. This indicates a tendency to overestimate the variable line.

3.—The six shorter variables are perceived a little more accurately at Rate 1 to Rate 4. The six longer lines are perceived best at Rate 3, and only slightly less accurately at Rate 4.

[60]Time and Accuracy of Judgment, Psy. Rev. 1911, 18, 189ff.

TABLE XIII

. Showing how many a's, b's, and c's were given by each observer for each comparison line; also a comparison of the % right for each line with the Accuracy record for the same line.

Rate 1

	95		97		98		98.5		99		99.5		100.5		101		101.5		102		103		105		Total	
	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c
Mz.	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	240	0
Dw.	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	240	0
Wn.	18	2	8	12	6	14	2	18	2	18	4	16	6	14	4	16	12	8	8	12	10	10	6	4	96	144
Ad.	8	12	8	12	14	6	16	4	14	6	10	10	14	6	10	10	10	10	12	8	12	8	12	8	140	100
Sc.	0	20	2	18	2	18	2	18	0	20	6	14	8	12	10	10	6	14	6	14	10	10	12	8	64	176
Ant.	12	8	6	14	6	14	8	12	6	14	8	12	2	18	6	14	14	6	8	12	12	8	6	14	97	146
Tot.	78	42	64	56	68	52	68	52	62	58	68	52	70	50	70	50	82	38	74	46	84	36	86	34	874	566
%R.	77.5		70		71.7		60		50.8		54.2		55		63.3		68.3		70.8		78.3		80.8		66.73	
%ab	65		53		57		57		52		57		58		58		68		62		70		72		60.7	

Rate 2

	95		97		98		98.5		99		99.5		100.5		101		101.5		102		103		105		Total	
	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c
Mz.	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	240	0
Dw.	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	240	0
Wn.	8	12	6	14	0	20	0	20	2	18	0	20	0	20	0	20	4	16	2	18	4	16	2	18	28	212
Ad.	14	6	10	10	10	10	20	0	8	12	10	10	10	10	10	10	12	8	16	4	16	4	18	2	154	86
Sc.	12	8	10	10	8	12	6	14	6	14	6	14	6	14	2	18	2	18	6	14	8	12	10	10	82	158
Ant.	10	10	2	18	2	18	8	12	4	16	2	18	0	20	2	18	4	16	6	14	6	14	2	18	48	192
Tot.	84	36	68	52	60	60	74	46	60	60	58	62	56	64	54	66	62	58	70	50	74	46	72	48	792	648
%R.	79.2		70.8		71.7		66.7		60.0		51.7		56.7		61.7		667		69.2		81.7		85.0		68.40	
%ab	70		57		50		62		50		48		47		45		52		58		62		60		55.0	

Rate 3

	95		97		98		98.5		99		99.5		100.5		101		101.5		102		103		105		Total	
	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c
Mz.	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	240	0
Dw.	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	240	0
Wn.	8	12	4	16	0	20	2	18	4	16	0	20	4	16	0	20	2	18	6	14	6	14	10	10	46	194
Ad.	14	6	16	4	14	6	14	6	14	6	14	6	14	6	12	8	8	12	16	4	16	4	16	4	168	72
Sc.	10	10	8	12	4	16	2	18	4	16	4	16	8	12	4	16	8	12	4	16	10	10	10	10	76	164
Ant.	12	8	4	16	6	14	6	14	8	12	8	12	8	12	4	16	2	18	2	18	6	14	12	8	78	162
	55		53		53		60		70				74	46	60	60	60	60	78	42	78	42	88	32	848	592
	56.7		53.3		65		76.7		81.7		%R.		55		68.3		71.7		76.7		86.7		91.7		70.28	
	66	54	70	50	64	56	72	48	84	36	Tot.		62		50		50		65		65		73		58.8	

Rate 4

	95		97		98		98.5		99		99.5		100.5		101		101.5		102		103		105		Total	
	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c
Mz.	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	240	0
Dw.	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	240	0
Wn.	6	14	6	14	0	20	2	18	0	20	0	20	4	16	2	18	0	20	2	18	4	16	4	16	32	208
Ad.	14	6	16	4	12	8	12	8	16	4	10	10	10	10	16	4	12	8	14	6	16	4	18	2	166	74
Sc.	2	18	4	16	16	4	4	16	0	20	2	18	2	18	6	14	8	12	8	12	1	28	10	10	74	166
Ant.	8	12	2	18	10	10	10	10	2	18	66	14	4	16	6	14	4	16	4	16	0	20	10	10	66	174
Tot.	70	50	68	52	78	42	68	52	58	62	58	62	60	60	70	50	64	56	68	52	72	48	84	36	818	622
%R.	78.3		72.5		68.3		68.3		56.7		48.3		45.8		71.7		75		71.7		83.3		82.5		69.4	
%ab	58		57		65		57		48		48		50		58		53		57		60		70		56.8	

Rate 5

	95		97		98		98.5		99		99.5		100.5		101		101.5		102		103		105		Total	
	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c
Mz.	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	240	0
Dw.	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	240	0
Wn.	14	6	0	20	2	18	14	6	6	14	2	18	6	14	0	20	4	16	4	16	2	18	6	14	60	180
Ad.	12	8	14	6	10	10	14	6	16	4	16	4	8	12	14	6	12	8	20	0	16	4	14	6	166	74
Sc.	4	16	0	20	2	18	4	16	0	20	4	16	6	14	6	14	6	14	2	18	10	10	8	12	52	188
Ant.	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	240
Tot.	70	50	64	66	54	66	72	48	62	58	62	58	60	60	60	60	62	58	66	54	68	52	68	52	758	682
%R.	84.2		68.3		63.3		64.2		51.7		49.2		46.7		58.3		66.7		73.3		80.8		84.2		65.9	
%ab	58		53		45		60		52		52		50		50		52		55		57		57		52.6	

Rate 6

	95		97		98		98.5		99		99.5		100.5		101		101.5		102		103		105		Total	
	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c	ab	c
Mz.	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	240
Dw.	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	240
Wn.	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	240
Ad.	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	240
Sc.	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	240
Ant.	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	240
Tot.	0	120	0	120	0	120	0	120	0	120	0	120	0	120	0	120	0	120	0	120	0	120	0	120	0	1440
%R	71.7		67.5		61.7		59.3		45		50		58.3		61.7		70.8		68.3		70.8		76.7		62.64	
%ab	0		0		0		0		0		0		0		0		0		0		0		0		0	

TABLE XIV

Percentage of Times each Degree of Confidence was Used.

A

Rates	1	2	3	4	5	6
Mz	78	100	93	100	100	0
Ad	10	10	20	13	10	–
Dw	87	93	90	100	97	100
Wn	5	1	1	–	2	–
Ant	22	17	10	10	–	–
Sc	–	–	–	–	–	–

B

Rates	1	2	3	4	5	6
Mz	22	–	–	–	–	–
Ad	50	60	20	56	60	–
Dw	13	7	10	–	3	–
Wn	35	11	18	13	23	–
Ant	18	3	20	17	20	–
Sc	27	34	36	30	20	–

C

Rates	1	2	3	4	5	6
Mz	–	–	7	–	–	100
Ad	40	30	60	31	30	100
Dw	–	–	–	–	–	100
Wn	60	88	81	87	75	100
Ant	60	80	70	73	80	100
Sc	73	66	64	70	80	100

4.—Confidence increases as the difference between the standard and variable increases. However, the degree of confidence given is largely an individual matter, and is not a very reliable index of objective accuracy. Observers range from those who report themselves as always sure, to those who never report themselves as sure.

5.—The individual differences are almost negligible. Except at Rate 1 the six observers are very close together.

3. Speed and Accuracy in Comparing Handwriting Specimens.

PROBLEM

Like the two preceding experiments the present experiment is concerned with the speed-accuracy relation in the perception of differences. As material for comparison I used specimens of handwriting, the specific problem being to discover, given a specimen of handwriting as a standard, how accurately other specimens can be compared with it, and a judgment of "better" or "worse" given. The speed, i. e. time allowed for perception and comparison, was varied by the experimenter and the accuracy was gauged by the number of correct comparisons.

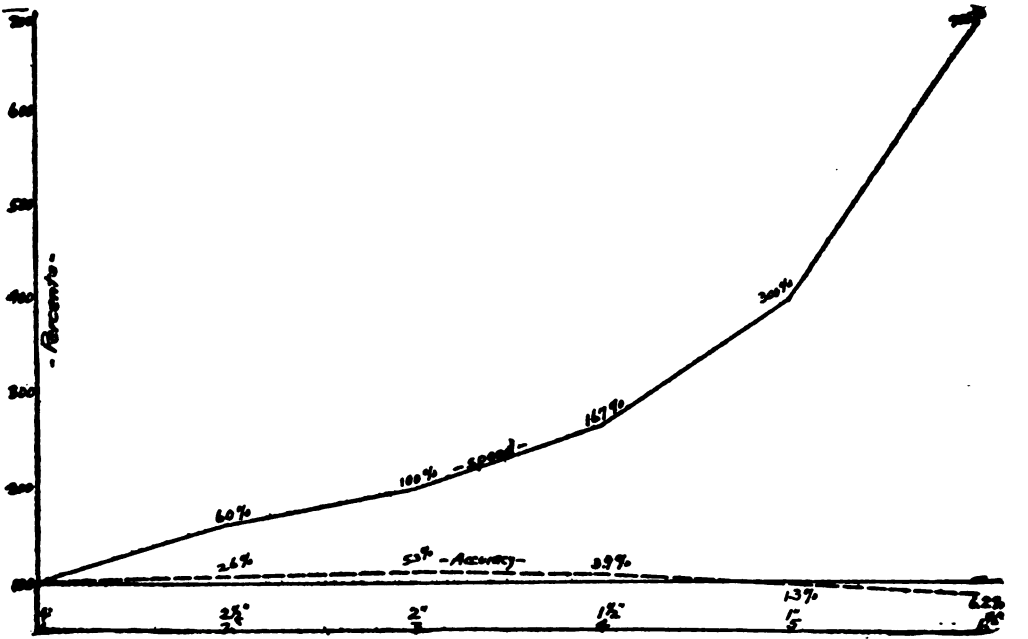
In such material as handwriting, the perceptions of differences which are capable of determining the direction of judgment is more difficult and more variable than in the case of weights and lines. For many factors such as general appearance of the writing, character, maturity, masculinity, size, shape, spacing of letters, idiosyncrasies in crossing t's, ending g's, y's, etc., regularity or "cramped" appearance of the characters, influence the decision: while the prejudice of the observer for some particular variety of handwriting complicates the process with a distinctly personal, and hence uncontrolled element. Thorndike,[81] apropos this subject, notes: "Being far, far more alike in sense organs and muscles than in central connections of neurones, we agree far better in comparing lines and weights than in comparing handwriting or poems."

It might seem from the foregoing considerations, that the choice between two specimens of handwriting is a purely personal affair, and is based on the individual's self-constituted standard alone. This is not strictly true, however, as the existence of handwriting scales shows that there are pretty well recognized norms in terms of which we judge handwriting as "good" or "bad;" and on which we all largely agree. Pintner[82] found, with 330 observers, a correlation of +.98 between 24 samples of handwriting graded by the Thorndike and the Ayres Scales. Thorndike[83] says that in using his scale, competent teachers without training will make an error of about .9 of a step in judging a sample, on the average, and that with practice judgments become more precise. Legibility, which was the criterion used by Ayres in constructing his scale, is probably the most important factor in determining the average person's norm for handwriting; though beauty, character, ease, etc., which Thorndike considered as well as legibility, certainly help determine it.

MATERIAL AND PROCEDURE

The material for this experiment consisted of 75 specimens of handwriting selected from nearly 200 printed specimens, all of which had been carefully graded by the Thorndike Scale. Samples which fell below 8 or went above 16 on this scale (the Scale ranges from 4 to 18 in steps of 1) were discarded, so that the specimens used might not be so far apart on the scale

[81] T. C. Record, 'Handwriting" 1910, *11*, p. 42 note.



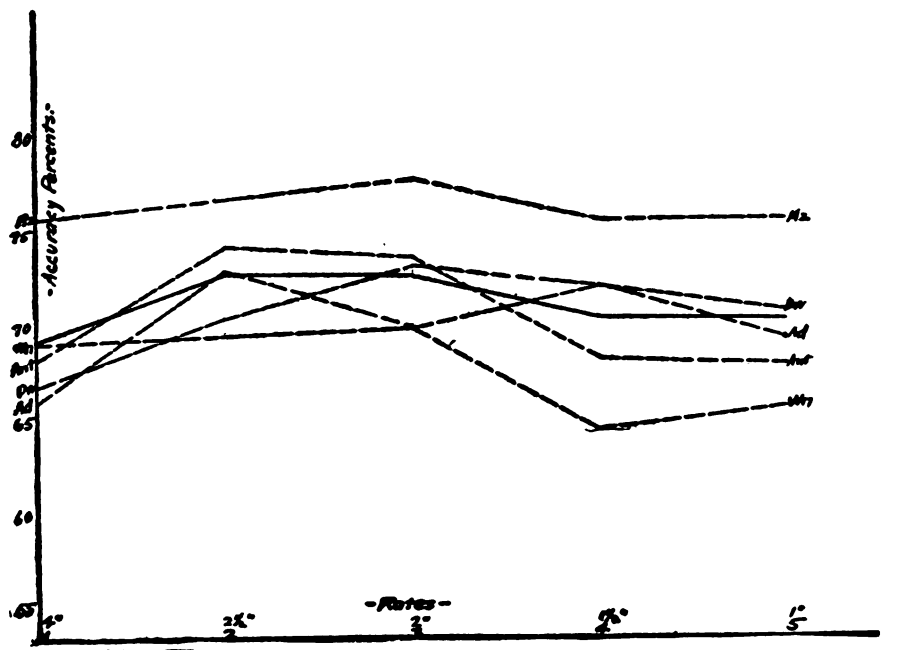


Diagram 7.—Showing relative changes in accuracy with increased speed. Rate 1 taken as 100 for both accuracy and speed.
Diagram 8.—Combined accuracy record for 5 O's and individual records for each. 5,250 comparisons at each rate.

as to be always recognized as different and hence make judgment too easy. Actually one specimen graded at 7.1 by the scale was included by accident, but this was the only sample below 8, and none above 16 were employed.

The 75 specimens were grouped into 5 sets of 15 each, the sets being as nearly as possible of equal value:—containing samples of approximately the same degree of excellence as measured by the Scale. Each specimen was mounted on cardboard 8″ by 2″ in size, and was numbered on the reverse side with two numbers, one for the group and one for the position in the group.

The method used in presenting the cards was that of Paired Comparisons. At the beginning of an experiment, the set of cards was placed before the subject, the cover card was removed, and card No. 1 was placed above and to the right of the other cards. The cards in the pack were then turned successively by the observer, exposed for a definite time, and a judgment of "better" or "worse" given in terms of the standard—in this case card No. 1. After the pack had been gone through, card No. 1 was replaced, and No. 2 became the standard: and so on until each card had been compared with every other one twice, once as standard and once as comparison. A statement of the degree of confidence in the judgments given was required, "a" sure, "b" fairly sure, and "c" uncertain, being used as in the previous experiments.

Five exposure intervals were employed, 4″, 2½″, 2″, 1½″, 1″,—the same intervals previously used—the rate being regulated by a bell metronome which rang on every fourth beat. The bell which was the signal to turn the next card, controlled the intervals with a high degree of exactness, and there was little confusion as to the length of time allowed for comparison.

The five sets of specimens were known as Set 1, 2, 3, 4, 5, respectively. Each set was presented at each rate according to the following plan of rotation: Rate 1, Set 1, Rate 2, Set 2, Rate 3, Set 3, Rate 4, Set 4, Rate 5, Set 5, so that in no instance did the same set or the same rate follow successively. In only a very few cases did an observer report that he recognized a

[8]A Comparison of the Ayres and Thorndike Scales, J. Ed. Psy. 1914, *5*, 525-526.

[9]Teachers Estimates of the Quality of Specimens of Handwriting. T. C. Record, Vol. 15, Nov., 1914.

TABLE XV

howing the number of Right Cases for each Observer and for each Rate:—also, the % or right cases, and the reliability of these averages, and the differences betwen them.

		Rate 1		Rate 2		Rate 3		Rate 4		Rate 5		
Ob.	Set	R	V	R	V	R	V	R	V	R	V	Tot.
Ms	1	130	15.3	105	47.9	123	30.0	114	33.7	120	27.6	592
	2	156	10.7	175	22.1	174	21.0	168	20.3	168	20.4	841
	3	174	28.7	177	24.1	178	25.0	180	32.3	168	20.4	877
	4	168	22.7	180	27.1	192	30.0	172	24.3	177	29.4	889
	5	168	20.3	171	18.1	150	3.0	159	11.3	162	14.4	810
Tot.		796		808		817		793		795		4009
%R		75.81		76.96		77.81		75.52		75.71		76.3
Ob.	Set	R	V	R	V	R	V	R	V	R	V	Tot.
Ad	1	122	23.3	147	5.9	120	33.0	126	21.7	104	43.6	619
	2	135	10.3	169	16.1	147	6.0	153	5.3	159	11.4	763
	3	168	22.7	162	9.1	180	27.0	168	20.3	168	20.4	846
	4	132	13.3	132	10.9	141	12.0	149	1.3	144	3.7	698
	5	135	10.3	156	3.1	144	9.0	159	11.3	153	5.4	745
Tot.		692		766		732		755		728		3671
%R		65.90		72.96		69.81		71.91		69.43		69.9
Ob.	Set	R	V	R	V	R	V	R	V	R	V	Tot.
Dw	1	139	6.3	109	43.9	145	8.0	131	16.7	136	11.6	660
	2	156	10.7	146	6.9	153	.0	159	11.3	156	8.4	770
	3	129	16.3	168	15.1	161	8.0	159	11.3	168	20.4	785
	4	132	13.3	165	12.1	150	3.0	154	6.3	126	20.6	727
	5	144	1.3	150	2.9	159	6.0	153	5.3	159	11.4	765
Tot.		700		738		768		756		745		3707
%R		66.67		70.29		73.14		72.0		70.95		70.6
Ob.	Set	R	V	R	V	R	V	R	V	R	V	Tot.
Wn	1	110	35.3	103	49.9	123	30.0	116	31.7	122	25.6	574
	2	149	3.7	155	2.1	150	3.0	137	10.7	146	1.6	737
	3	151	5.7	149	3.9	150	3.0	142	5.7	118	29.6	710
	4	179	33.7	174	21.1	163	10.0	167	20.7	161	13.4	844
	5	137	8.3	150	2.9	148	5.0	113	34.7	141	6.6	689
Tot.		726		731		734		675		688		3554
%R		69.14		69.62		69.90		64.29		65.52		67.7
Ob.	Set	R	V	R	V	R	V	R	V	R	V	Tot.
Ant	1	123	22.3	125	27.9	132	21.0	115	32.7	112	35.6	607
	2	151	5.7	162	9.1	150	3.0	136	11.7	143	4.6	742
	3	156	10.7	173	20.1	179	26.0	162	14.3	162	14.4	832
	4	136	9.3	155	2.1	143	10.0	160	12.3	138	9.6	732
	5	152	6.7	166	13.1	170	1.7	142	5.7	159	11.4	789
Tot.		718		781		774		715		714		3702
%R		68.38		74.38		73.71		68.09		68.0		70.5
Sum		3632		3824		3825		3694		3690		
Av.		145.3		152.9		153.0		147.7		147.6		
%R	(total)	69.18		72.84		72.86		70.36		70.29		
PE	(Av)	1.11		1.39		1.17		1.24		1.27		
Dif		3.68		.02		—		2.50		2.55		
in Avs.												
PE dif		1.61		1.82		—		1.70		1.73		

specimen which he had seen at a some previous rate; this usually happened with a very poor or a very good specimen, or with one distinctive in some respect. I did not think that these cases were numerous or noteworthy enough to affect the general result, and therefore, no account, (other than a mention of them in the introspections) is taken of them.

PRACTICE AND FATIGUE

The use of five groups was chiefly for the purpose of preventing the necessity of presenting all the specimens at one rate successively. This precedure, if used, would very probably have introduced a memory or practice factor, and the last rate used would have enjoyed a great advantage over the others. Alternation of rates and sets should make the reliability of the general result very high; especially since the experiments extended over a considerable time, with intervals varying from a day to several weeks between, and included 26,250 comparisons in all. Rest intervals, and the relatively easy character of the experimental work ruled out the factor of fatigue.

RESULTS

Of the six men who acted as subjects in the preceding experiments, one, Sc, was unable to serve in the present experiment. He was forced through illness to leave college, and his record could not be secured. The results are, therefore, based on the judgments of five observers.

Table XV incorporates the results of the experiments. As each subject made 1050 comparisons at each rate, the sum for any one rate, eg. 3632 right cases, represents the number of right judgments in a total of 5250 trials. The average number of right cases at a given rate, means the average number of right judgments out of 210 trials, and is based on 25 measures, 5 for each observer. The % right cases for each rate, with the PE (Av.) are also given; in every instance the differences of the averages, and the PE's (diffs.) have been reckoned from Rate 3, as the largest number of right cases appeared at this rate. Diagram 8 gives, graphically, the facts of Table XV. In this diagram the rates are laid off at regular intervals on the baseline, and the accuracy records are plotted in % R-cases on the Y-axis. The total accuracy curve rises from Rate 1 to Rate 2, and remains on a level from Rate 2 to Rate 3. The accuracy falls off slightly at Rate 4, and continues

on the same reduced level to Rate 5. It will be seen that there is no difference between Rates 2 and 3, and practically none between Rates 4 and 5. However, the chances are 16 in 17 of a real increase in accuracy from Rate 1 to Rate 2, and 4 in 5 of a loss in accuracy from Rate 3 to Rate 4.

Diagram 9 shows graphically the corresponding changes in speed (reduced exposure rate) and accuracy (correct comparisons). Both factors are measured from Rate 1 as unit or 100%. The general appearance of the curves resembles closely the relation previously found between the two factors in the case of weights and lines. Starting with Rate 1, an increase in speed of 60% (i.e. decrease in exposure interval from 4″ to 2½″) caused an increase in accuracy of 5.3%, which remained constant when the speed increased 100%; an increase of 167%, and even 300%, in rate of exposure produced only a relatively slight decrease from this level, the accuracy curve still remaining 1.7% above the level (100%) of Rate 1. It appears that specimens are compared as accurately when 1″ is allowed as when 4″ are given: the speed may be increased 300% with no loss in accuracy. Speed and accuracy are again directly related up to a certain point, after which they are inversely related.

Before attempting to explain the behavior of the accuracy curve in Diagram 9, it might be well to consider the introspections of the subjects. At the conclusion of the experiment, each subject was asked to state the basis of his judgment as far as possible, and to say what effect the increased speed had on his confidence in his answers. Specific replies were requested.

INTROSPECTIONS

Mz.—I based my judgments on legibility, principally,—spacing, uniformity of lines, heaviness, etc. I prefer a medium slant over a vertical. Rate 1 allowed too much time; there was a tendency for attention to wander, and I sometimes doubted the judgment I had just given. I preferred Rates 4 and 5. At these rates I usually took a good look at the standard, and made comparisons in terms of this visual image, without troubling to look back at the standard again. I was always "a" confident, that is, I felt that I was giving the best judgment of which I was capable. I don't remember recognizing any specimen particularly.

Ad.—I judged as "better" those specimens in which the writing was regular, smooth, flowing. Rate 3 was by far the best rate; it allowed time enough for comparison without any surplus time, or feeling of being hurried. At the first rate, I believe that the long interval allowed attention to wander and permitted irrelevant details—irritation, etc., to come in. I compared the specimens directly. I recognized two samples:—one a very fine sample of slant writing, and the other a heavy vertical. I don't think that this influenced my comparisons.

Dw.—I based my judgments on legibility; formation of letters and general appearance. I prefer a slant. Rate 4 seemed to be the best interval for me, as the slow rates became very tiresome. I tried to compare each pair of specimens directly, though I often felt that my judgment was influenced somewhat by my standard of "good." So far as I know, memory of previously given comparisons did not affect my answers.

Wn.—My judgments were based on the general appearance of the handwriting, spacing, smoothness, etc. Rate 1 was too long, and I often found myself reading the sample through, rather than examining the writing. The other rates were about equally as good. Once or twice I found my judgment influenced by the resemblance of the sample to the handwriting of a friend. I also recognized the tendency to give the same degree of confidence that I had previously given. I remembered several of the specimens rather indefinitely but I don't think that it influenced my judgment.

Ant.—I preferred those specimens which were written smoothly, and regularly. I think that Rates 2 or 3 gave about the best balance, not too fast nor too slow. I compared the specimens directly; I think that I recognized a few, though I am not very certain. I don't believe that memory affected my judgment.

The explanation of the relationship of accuracy to the time of exposure in the case of handwriting is very probably similar to the explanation given for the relationship in the experiments on lines and weights. The situation in the three experiments are not esentially different, except that in the two former experiments accuracy of comparison depended primarily on native sensitivity, while in the present experiment the norm of comparison is largely the product of education and training.

In none of the experiments was there any need,—or much opportunity,—for factors other than those contained directly in the situation entering into the comparison. During the course of the experiment, the observations of the experimenter as well as the introspections of the subjects indicated that Rate 1 was too slow. Comparatively a large part of the interval was unused, judgments were often given very quickly, and during the remainder of the interval the observer merely waited for the bell to signal the next comparison. Slight irritation, lapses of attention, boredom, were often in evidence. Rates 2, 3, and even 4 seemed to be better balanced; i.e. they allowed time for the cards to be turned and judgment to be given in a regular and rhythmic fashion. Rate 5, while a bit strenuous both for the observer and the recorder, still allowed time enough for the observer to register his comparison. I found it impossible for the observer to vocalize, or the recorder to note down comparisons, when less than 1″ was allowed. It might seem that the manipulation, eg. turning of the cards, was the main reason for the loss in accuracy at Rates 4 and 5, but the observers became so skillful in turning the cards, and there were so few interruptions due to a loss of rhythm, that I do not believe that this factor played any considerable part in determining the accuracy. Also, the lowest accuracy record was at Rate 1.

Rates 2 and 3 insofar as they deserve to be called 'optimal" derive their superiority from the fact that they gave just time enough for the factors determining the judgment to mature, more specifically, for the observer to perceive both stand-

TABLE XVI

Showing the Per Cent of "a" and "b" Judgments given by each Observer at each rate, compared with the Per Cent of right cases at same rate. Five Observers,—5250 comparisons by each.

	Rate 1 Per Ct.		Rate 2 Per Ct.		Rate 3 Per Ct.		Rate 4 Per Ct.		Rate 5 Per Ct.		Totals Per Ct.	
Obs.	a b	R	a b	R	a b	R	a b	R	a b	R	a b	R
Mz.	100	75.8	100	77.0	100	77.8	100	75.5	100	75.7	100	76.3
Ad.	60	66.0	56.2	73.0	61.4	69.7	61.8	72.0	63.3	69.4	60.6	69.9
Dw.	80	66.7	79	70.3	66.7	73.1	75.2	72.0	80.5	71.0	76.0	70.6
Ant.	71.8	68.4	74.3	74.4	73.8	73.7	60.0	68.0	73.8	68.0	70.6	70.5
Wn.	64	69.2	59.5	69.6	59.0	70.0	55.2	64.2	59.5	65.5	59.4	67.7
Per Ct. R	74.7	69.2	73.8	72.8	72.2	72.9	70.4	70.4	75.4	70.3	73.5	71.0

ard and comparison clearly enough and long enough to isolate those details, shape, legibility, appearance or what not through which the comparison was made. A longer interval, therefore, must have allowed irrelevant facts to slip in; while

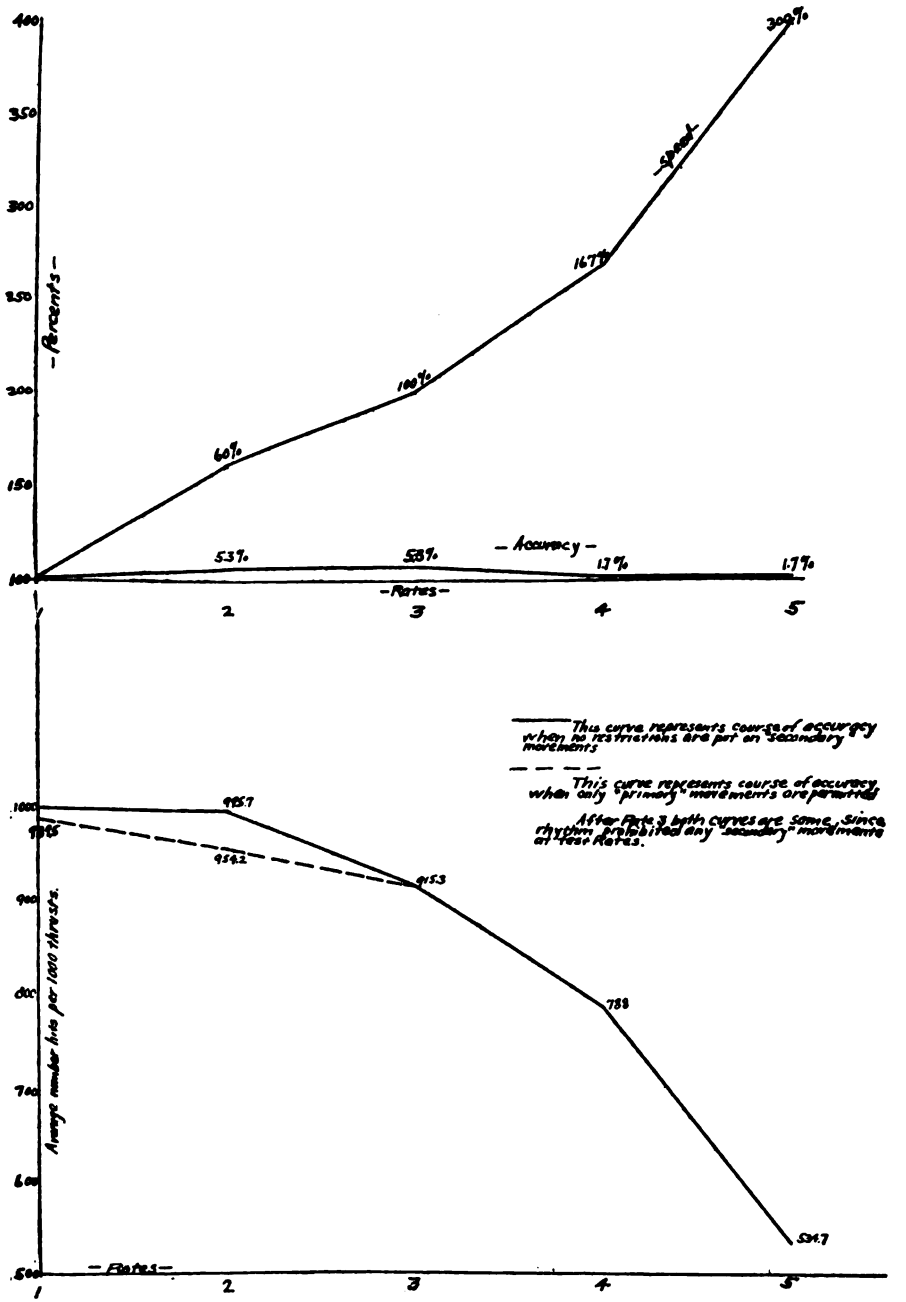


Diagram 9.—Showing the corresponding changes in speed and accuracy measured from Rate 1 as 100.

Diagram 10.—Showing the number of hits per 100 thrusts at each rate.

a shorter interval excluded certain necessary data for accurate comparison. Both conditions lead to inaccuracy.

Table XVI gives the % of " a and b," sure and fairly sure, judgments returned by the observers, compared with the % of right judgments at each rate. The confidence record shows little change from Rate 1 to Rate 5, in fact, the % a's and b's is nearly the same as the % accurate judgments. On the average, the observers are as confident at one rate as another. Individual preferences for one or more categories as shown in the previous experiments, are followed very closely in the present experiment. Mz was equally confident throughout; Ad gave about 60% of a's and b's, no matter what the rate. Dw and Ant are confident about as often as they are accurate, while Wn gave a majority of a's and b's, though, as in previous experiments, he gave a large number of c's also. On the whole, the confidence record is here a pretty accurate index of the correctness of the observer's answers. One is justified, in view of the above results, in concluding that accuracy in these comparisons was usually correlated with confidence; though the reverse may or may not be true.

A word should be said in regard to the relative difficulty of the sets of handwriting employed. As was remarked before, the sets were intended to be as nearly as possible of equal difficulty, as determined by scale. To test them, I added up the 210 differences between the 15 specimens in each set with the following results: the sum of the differences for Set 1 was 537; for Set 2, 673.9; for Set 3, 506.4; for Set 4, 626.3; for Set 5, 588.5. Since the larger the gap between two specimens, the easier the judgment, it would seem at first glance that Set 2 was the easiest, and Sets 4, 5, 1, and 3 followed in the order given. This conclusion is not tenable, however, since a few large differences might be compensated for by many small (and hence, difficult) differences. Table XVI shows that Set 1 was the most difficult, Set 5 next, then Sets 3, 2, and 4, in order. The difference in the number of right comparisons was much greater between Sets 1 and 5, than between Sets 2, 3, 4, or 5.

CONCLUSIONS

The conclusions to be drawn from this experiment as to the relation of speed and accuracy are as follows:

1.—In the present experiment, speed and accuracy are both

directly and inversely related. As the rate increases, eg. the interval for judgment decreases, from 4″ to 2½″, there is a slight but appreciable rise in the accuracy of comparison. From 2½″ to 2″ the accuracy record does not change, but at Rate 4, eg. 1½″, there is a small loss in accuracy which does not increase or diminish when the interval becomes 1″. The accuracy record at 1″, is however, higher than at 4″.

2.—The optimal interval is at 2½″ or 2″.

3.—The degree of confidence, as expressed by the terms, *a*, *b*, *c*, does not change appreciably from the slowest to the fastest rate. The % of "sure" and "fairly sure" judgments closely parallels the accuracy records.

4.—In the use of the confidence categories, observers retain the preferences shown in the preceding experiments. One observer gave 100% "a" judgments, while one gave 60% "a" and "b" answers. Evidently, the terms expressing confidence are interpreted differently.

SECTION III

COORDINATION EXPERIMENTS.

1. Thrusting Experiment. 2. Maze Tracing Experiments.

PROBLEM

These two experiments were planned for the purpose of tracing the course of accuracy for a fairly simple series of muscular movements when the speed with which the movements are made is varied in a definite and regular manner. In both experiments the speed was the variable factor, and the accuracy for each "Rate" was measured in terms of "misses" or "touches." No judgment is required, the task being purely one of voluntary muscular coordination in which the reagent simply makes the prescribed movements at the dictated rates.

APPARATUS

The Three Hole Aiming Test, which was used for the Thrusting Experiment, consists of a wooden board on which is fastened a wooden triangle inclined to the board at an angle of 45 degrees. The triangle, which is equilateral, has a brass-lined hole in each angle, and the subject's task is to thrust a metal rod or stylus into the three holes alternately. The stylus is just large enough to fit into the holes easily. Each successful thrust into a hole is recorded on a clock which is in electrical connection with the brass-lined holes and the metal stylus. This apparatus was designed by Prof. Woodworth and used by him at the World's Fair in 1904. It has been used by Stecher[84] and Hollingworth[85] as a test of efficiency in motor co-ordination under different physical or physiological conditions. As a measure of voluntary muscular control, this test is very useful, as it requires coordination of a group of muscles under the guidance of visual impression, i.e. accurate co-ordination of eye and arm in the performance of a definite task.

The ease with which the speed can be controlled (i.e. rate of thrusting) and the accuracy (hits) measured, recommend-

[84]Effects of Humidity on General Eciency, Archives of Psy. *38*, 1916.
[85]The Influence of Caffeine on Mental and Motor Efficiency, Archives of Psy. *22*, 1911.

ed this apparatus as useful in a study of the speed-accuracy relation. It is true that there is no way of measuring the amount of error or "miss" except in lengthened time, but for the purposes of this experiment it seemed sufficient simply to take all errors as of equal value. This procedure does not, of course, take into account the amount of "miss," but the range for misses is so small that I did not think it worth considering.

The Three Hole Aiming Test is essentially a target test, in which a rapid thrusting movement is required. As a test of a motor coordination of a slightly different kind I have used a maze test in which the movement is a continuous one—more like a writing movement—and in which steadiness is an important factor. The Coordination Test (Johnson-Dunlap) of the Johns Hopkins Series was used, since this test is prepared for tracing regulated by a metronome. Tracing is done from right to left, so that the pathway ahead may not be hidden by the hand. One stroke is made to each beat of the metronome, and accuracy is measured by the number of times the traced line touches the guide lines. This particular maze test consists of 6 sections in duplicate, and as each section requires 17 separate movements, there are 102 movements made in tracing one blank.

PROCEDURE

The procedure followed in the thrusting experiment was very simple. With the subject seated before the board, the metronome was started and the subject was instructed to thrust the stylus into the three holes in order, one thrust for each beat. The first thrust was in the lower left hand hole, and the order of thrusting was counter-clockwise. Five different rates of thrusting were used, with the metronome set to beat 60, 96, 120, 160, 240 (approx.) times per minute. One thrust, therefore, occupied 1″, ⅝″, ½″, ⅜″, or ¾″. One hundred thrusts at any one rate constituted one trial. To insure one thrust for each beat of the metronome, and also 100 and no more thrusts—the experimenter counted in time with the metronome, breaking the electrical circuit on the 100th count. At the end of each trial, the number of hits were read from the clock.

Before beginning a trial at a changed rate, a preliminary series of thrusts was taken in order to accustom the subject

to the new rhythm and thus avoid any sudden "jump" in the accuracy of the succeeding trials due to poor adjustment in the first. This plan was also intended to cut down the interference effect which might arise in the shift from one rate to another. In spite of these "adjustment" series, the practice effect was pronouncedly present at all rates as may be seen by referring to the succeeding trials for any one rate. The trials were taken according to a regular plan, eg. 100 thrusts at Rate 1, 100 at Rate 3, 100 at Rate 5, etc. This was done in order to spread the practice effect equally, and not give any one rate the advantage in regard to practice. Two or three minutes rest were allowed between trials, and never more than 1000 thrusts were taken at one sitting;—hence fatigue was not considered as a factor in the experiment.

In the tracing experiment, the subject was instructed to start at the cross marking the entrance of the upper right hand pathway, and make one move for each beat of the metronome. A fountain pen was used for the tracing, as a pencil changes the width of the line between the start-point and the finish. After tracing the first pathway, the subject returns to the pathway immediately below, pausing for a beat or two in order to catch the rhythm. This interval between pathways varied slightly according to the rate,—usually occupying a second or two. No attempt was made to control it, other than to keep it from becoming too long or too short. Ten sheets were traced by each subject at each rate. Since each sheet required 102 movements, as stated before, 1020 movements at each rate or 5100 movements in all were made by each subject.

The time intervals used for tracing were the same as those used in the thrusting experiment. With the metronome set at 60, 96, 120, 160, 240, each movement occupied 1″, ⅝″, ½″, ⅜″, ¾″. A regular order of trials was followed:—no two sheets were traced successively at the same rate. This plan, as in the thrusting experiment, was designed to eliminate a concentration of practice effect at one rate. The whole experiment did not occupy, for one subject, more than two hours; and frequent rest intervals are believed to have ruled out fatigue as a factor.

1. *Thrusting Experiment*

RESULTS

In performing the thrusting experiments, the data for which are to be found in Table XVII, the subjects were given no specific instructions as to the method to be used in making the thrusts. They were simply told to keep the rhythm,—make one thrust to every beat of the metronome. At Rate 1, the slowness of the rhythm permitted a subject to correct an inaccurate thrust or "primary movement" by a later small "secondary movement"; for example, if the stylus missed the hole by a small margin, it could be quickly pushed in and a hit recorded without losing the rhythm. This use of a "secondary movement" was noticeable, though to a somewhat lesser extent at Rate 2 also. At the faster rates, 3, 4, 5, the speed of the thrusting movement precluded any later adjustments, without sacrificing the rhythm, and, in consequence, the accuracy depended entirely on the initial thrust or "primary movement." In order to make the accuracy records comparable for all rates, I conducted a second set of experiments at Rates 1 and 2 only, in which the subject was instructed NOT to make any corrections of an inaccurate thrust no matter how much time he had for doing so. In a sense, the first method is the most logical one, since, from a practical standpoint, the slower rates should be given the advantage which they offer of later adjustments,—if for no other reason than as a compensation for their slowness. In presenting the results of the experi-

TABLE XVII

Records (errors) 6 Subjects:—5 Rates, 30,000 Thrusts.

An error is defined as a "miss." Later adjustments, eg. in case of a miss, not prohibited;—allowed as long as Subject did not lose rhythm. These later adjustments immediately after a thrust were used generally at Rates 1 and 2, but speed of thrusts did not permit them at Rates 3, 4, or 5.

	Rate 1		Met. 60			100 thrusts in 100"						
Trials	1	2	3	4	5	6	7	8	9	10	sum	hits	dev.
Errors													
Mz.	0	0	0	0	0	0	0	0	0	0	0	1000	–
Ad	0	0	0	0	0	0	0	0	0	0	0	1000	–
Dw	0	0	0	0	0	0	0	0	0	0	0	1000	–
Wn	0	0	0	0	0	0	0	0	0	0	0	1000	–
Ant	0	0	0	0	0	0	0	0	0	0	0	1000	–
Sc	0	0	0	0	0	0	0	0	0	0	0	1000	–
Total	0	0	0	0	0	0	0	0	0	0	0	6000	–
Av.			0				0					1000	
Acc%			100				100					100%	–

	Rate 2			Met. 96			100 thrusts in 62.5"						
Trials	1	2	3	4	5	6	7	8	9	10	sum	hits	dev.
Errors													
Mz.	4	2	0	0	0	0	0	0	0	0	6	994	1.7
Ad	3	2	0	0	1	0	0	0	0	0	6	994	1.7
Dw	2	2	0	0	0	0	0	0	0	0	4	996	.3
Wn	0	0	0	0	0	0	0	0	0	0	0	1000	4.3
Ant	5	0	1	2	0	0	0	0	0	0	8	992	3.7
Sc	1	0	1	0	0	0	0	0	0	0	2	998	2.3
Total	15	6	2	2	1	0	0	0	0	0	26	5974	14.0
Av.	2.5	1.0	.3	.3	.2	0	0	0	0	0	4.3	995.7	
A.D.	1.1	.5	.2	.3	.1	0	0	0	0	0	1.6	2.3	
Acc%			99.1				100					99.6%	

	Rate 3			Met. 120			100 thrusts 50"						
Trials	1	2	3	4	5	6	7	8	9	10	sum	hits	dev.
Errors													
Mz	20	11	8	5	4	6	4	3	4	6	71	929	13.7
Ad	24	22	19	7	8	12	9	10	9	7	127	873	42.3
Dw	22	21	15	4	2	2	1	2	2	2	73	927	11.7
Wn	6	6	3	3	3	4	4	4	3	1	37	963	47.7
Ant	18	16	12	15	18	20	10	15	9	10	143	857	58.3
Sc	12	8	4	10	3	1	8	3	3	5	57	943	27.7
Total	102	84	61	44	38	45	36	37	30	31	508	5492	201.4
Av	17	14	10.2	7.3	6.3	7.5	6.0	6.2	5	5.2	84.7	915.3	
A.D	5.3	5.7	5	3.4	4.4	6.2	3	4.2	2.7	2.5	33.4	33.6	
Acc%			89.0					94.0				91.5%	

	Rate 4			Met. 160			100 thrusts 37.5"						
Trials	1	2	3	4	5	6	7	8	9	10	sum	hits	dev.
Errors													
Mz	30	35	23	19	18	19	25	23	11	22	225	775	3.0
Ad	24	28	25	27	24	18	27	19	22	28	242	758	30.0
Dw	49	45	29	31	16	14	20	23	18	16	261	739	49.0
Wn	15	15	18	19	16	15	11	11	11	11	142	858	70.0
Ant	30	36	27	31	20	22	15	20	19	18	238	762	26.0
Sc	28	23	20	19	11	7	23	15	9	9	164	836	48.0
Total	176	182	142	146	105	95	121	111	90	104	1272	4728	226.0
Av.	29.3	30.3	23.7	24.3	17.5	15.8	20.2	18.5	15	17.3	212	788.0	
A.D.	7.0	8.3	3.3	5.3	3.2	4.0	4.5	3.7	4.7	5.0	35.3	37.7	
Acc%			75%					82.6%				78.8%	

	Rate 5			Met. 240			100 thrusts 25"						
Trials	1	2	3	4	5	6	7	8	9	10	sum	hits	dev.
Errors													
Mz	50	46	59	49	53	42	46	51	51	50	497	503	31.7
Ad	58	53	56	47	58	44	43	49	42	40	500	500	34.7
Dw	67	60	53	46	40	43	43	39	37	35	463	537	2.3
Wn	44	42	45	46	40	37	42	33	46	45	420	580	45.3
Ant	47	53	51	42	47	50	43	49	45	40	467	533	1.7
Sc	53	44	48	39	40	54	46	42	39	40	445	555	20.3
Total	319	298	312	269	278	270	263	263	260	260	2792	3208	136.0
Av.	53.2	49.7	52	44.8	46.3	45	44	44	43.3	43.3	465.3	534.7	
A.D.	6.0	5.7	4	2.3	6.3	4.7	1.5	6	4.0	5.0		22.7	
Acc%			50.8				56.1					53.5	

ment, I have considered first the results of the trials in which the subject was left free to use "secondary movements" as long as the rhythm permitted;—and secondly, those trials in which no correction of an inaccurate thrust was allowed.

Table XVII incorporates the results of 30,000 thrusts, made by 6 subjects at 5 rates of thrusting. Errors for each trial, as well as the errors and hits for all 10 trials, are given, for each individual and for all 6 individuals taken together. Diagram 10 shows graphically the course of the error curve at the different rates. The diagram is purely schematic,—no account being taken of the increases in speed; the rates are represented merely by equal intervals on the base line.

Probably the first impression which one gets from Diagram 10 and Table XVII is the apparently very rapid "drop in accuracy, especially the decrease from Rate 3 to Rate 5. This loss in accuracy is not, however, as great as it seems at first glance;—nor as the curve would indicate. For we have yet to consider the changes in rate of thrusting which run parallel to these changes in accuracy. This is done in Diagram 11. Here the rate of thrusting is taken as 100:—Rate 2 is then 60% greater, Rate 3 100% greater, Rate 4 167% greater, and Rate 5 300% greater, than Rate 1. The accuracy at Rate 1 is also taken as 100%. The loss in accuracy at Rate 2 is then only .4 of 1% from Rate 1; at Rate 3 the loss is 8.4%, at Rate 4, 21.2%, and at Rate 5, 46.5%. Thus it is apparent, that, while Speed and Accuracy are inversely related, the decrease in accuracy at a given rate does not compare with the increase in speed at the same rate. In fact, the loss in accuracy as far as Rate 3 is negligible, when misses are considered in relation to hits. The representation of changes in accuracy by errors alone has led to many false conclusions as to diminution in efficiency due to fatigue, changes in conditions, changes in procedure, etc. Thorndike[86] after citing the results of Sikorski, who in testing the same children before school and after school in writing from dictation, found the average per cent of wrong letters to be 33⅓% greater in the late tests, says: "I trust that the reader is not so unsophisticated as to assume that the above figures, even if taken at face value, show an efficiency before school of 1⅓ times that existing after school. They as truly mean that, since about 99.3% of the letters were correct in the morning and about 99% after school, the efficiency before school was 1.0033 times that existing after

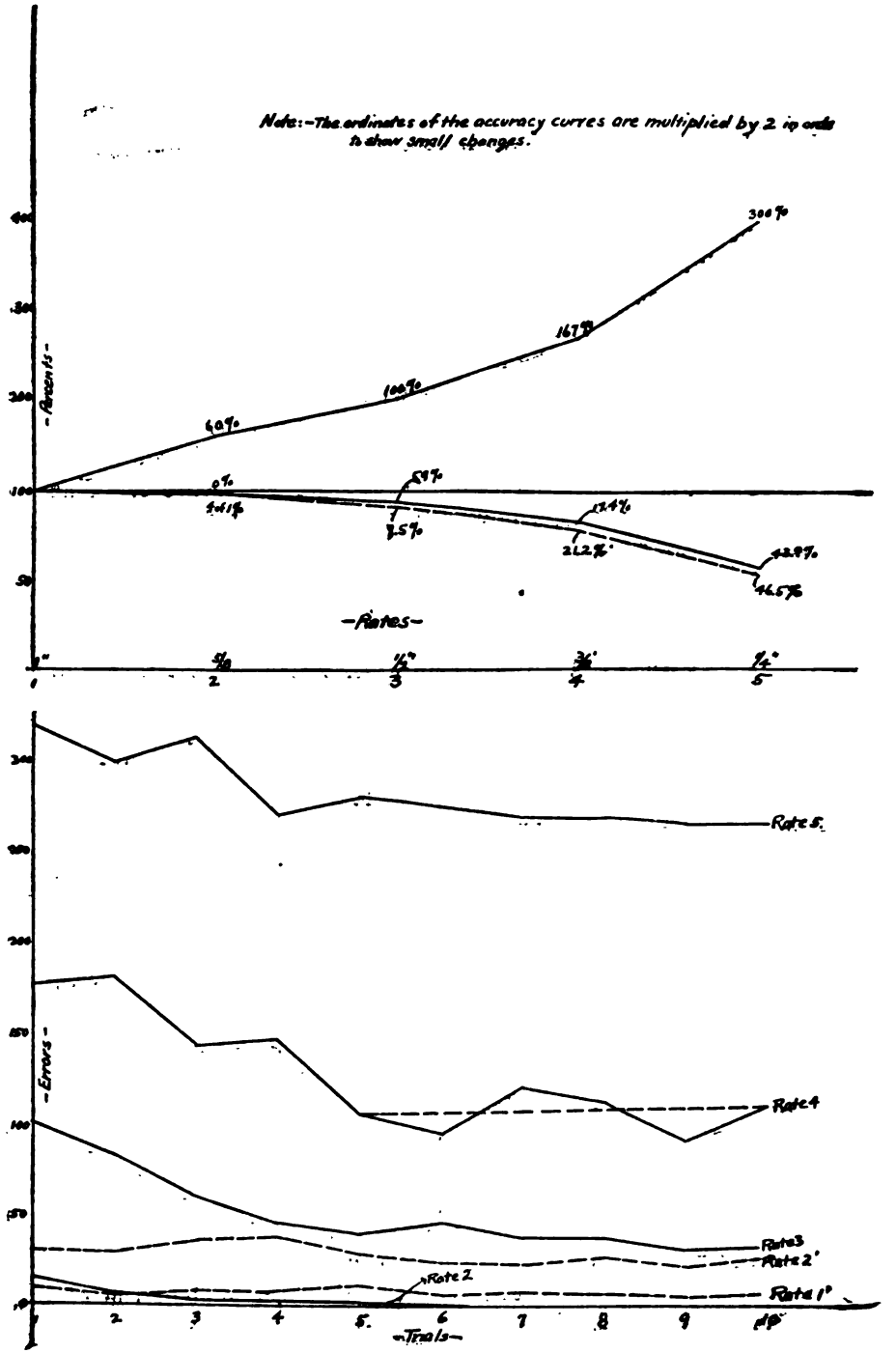


Diagram 11.—Showing the changes in the accuracy of thrusting corresponding to the changes in rate of thrusting. Both speed and accuracy measured from Rate 1 as 100. The solid line represents the accuracy records from the second half of the experiment

school." In thus minimizing the loss in accuracy, in relation to the increase in speed, it must not be overlooked that we could hardly expect the accuracy to vary equally with the speed. In the present experiment, for instance, accuracy can only fall from 100% to 0, while the speed can increase almost indefinitely.

Table XVII contains another fact, just as obvious as the speed-accuracy relation, and much more real on the face of it: namely, the very evident improvement shown by all the subjects, no matter what the rate. This improvement is largely attributable to practice. Diagram 12 shows the changes in accuracy (in terms of errors) for each trial from 1 to 10. The practice curves for each rate, with the exception of Rate 1, are plotted separately. At Rate 2 improvement ends at trial 3, the errors drop out completely at trial 6, and do not appear again; only trials 1 and 2 give evidence of appreciable error and some of these errors are almost certainly due to the awkwardness and lack of adjustment, uncertainty, etc., nearly always present at the beginning of any experiment. In spite of the preliminary series given with the express intention of eliminating just such errors, some subjects are slower than others in "warming up," and some become more easily confused and "thrown off the track." Hollingworth[87] in speaking of this test (3 Hole) remarks that skill in the test varies for quite unaccountable reasons, and that much seemed to depend on beginning luckily and striking a favorable rhythm. The fluctuations in accuracy for thrusts at Rate 3, are very slight after trial 5; the flatness of the curve from this point on indicating that practice has practically spent itself, though it is very clearly present in the first four trials. Rate 4 shows rather erratic changes in accuracy. The decrease in errors gives evidence of practice during the first 5 trials; but from trial 5 to trial 10, the accuracy, for some unknown reason, fluctuates up and down, though the changes are not so great, nor so evidently in the one direction of improvement, as in the first 5 trials. A straight line drawn from a point representing trial 5 to the point representing trial 10, by balancing the factors that make for accuracy, i.e. practice, adaptation, etc., vs. those that make for inaccuracy, eg. speed, loss of rhythm, etc., would probably give a fairer picture of the accuracy curve in this region. The changes in

[87]Ed. Psy. Vol. III, p. 81 ff.

the curve for Rate 5 are practically all in the first 5 or 6 trials; after trial 5 the flatness of the curve indicates that no further significant changes from practice are to be expected.

The trend of the curves in Diagram 12 seems to warrant the assumption that practice is negligible as a factor in improving accuracy after the first five trials. Accordingly, I have divided the thrusts at each rate into two groups (Table XVII) the first comprising trials 1 to 5, and the second trials 6 to 10. Each group represents, therefore, the records of 3,000 thrusts—the second, with practice eliminated.

The accuracy at each rate when corrected for practice, is represented in Diagram 11. It is worthy of note, as showing the effect of practice, that the decrease in accuracy for Rates 2, 3, 4, 5, measured from Rate 1 as 100%, is at every rate less than when all the trials are taken into account. Except at Rates 1 and 2, the elimination of practice means about 3% increase in the accuracy record at each rate.

In examining the factors that enter into an accurate thrust, I have used the terms, "primary movement" and "secondary movement." These terms are used, as was said before, to represent the initial thrusting movement, and the later small movements which are necessary for a hit when the subject's aim is inaccurate. It was evident to the experimenter that a large share of the credit for the high accuracy at the first two rates must be given to this second factor, which at Rates 3, 4, 5, the speed of the rhythm ruled out. A hit at these rates depended almost entirely on the accuracy of the initial aim.

This conclusion is substantiated by the second series of experiments, and also by the more detailed analysis of a voluntary movement made by Woodworth.[88] In separating out the factors involved in a voluntary movement, Woodworth distinguishes between two chief influences, the "initial adjustment," and the "current control." The operation of these factors can best be shown by an example. In drawing lines on a revolving drum, with the requirement that each line equal that which immediately precedes it, the "initial adjustment" or preliminary adjustment, of the movement is to be distinguished from the "current control" of the movement

[87]The influence of Caffeine on Mental and Motor Efficiency, Archiv. of Psych. 22, p. 49.

[88]Op. cit. 27-62.

process, eg. the drawing of the line itself, and the fine adjustments at the end of the line. An increase in the speed of the movement with a consequent decrease in the interval between successive movements affects these two factors differently. The shortened interval aids initial adjustment by setting up a "momentum of uniformity," eg. making the movements regular, rhythmic, and automatic, while the increased speed works for inaccuracy by making impossible

TABLE XVIII

6,000 thrusts at two Rates 1 and 2: 6 Subjects: Records in terms of "misses." Subjects were instructed not to use secondary movements. If a thrust missed a hole it was counted a miss, no correction being allowed.

Rate 1

Trial Error	1	2	3	4	5	6	7	8	9	10	sum	hits	dev.
Mz	2	2	1	0	0	1	2	1	1	0	10	990	.5
Ad	1	2	1	4	5	2	1	3	1	2	22	978	11.5
Dw	2	0	2	1	3	0	1	0	2	1	12	988	1.5
Wn	0	0	0	0	0	0	0	0	0	0	0	1000	10.5
Ant	4	1	2	1	2	0	1	2	0	3	16	984	5.5
Sc	1	0	1	0	0	1	0	0	0	0	3	997	7.5
Total	10	5	7	6	10	4	5	6	4	6	63	5937	37.0
Av.	1.7	.8	1.2	1	1.7	.7	.8	1	.7	1	10.5	989.5	
Ad.	.7	.5	.4	.5	.8	.3	.3	.5	.3	.5		6.2	
%acc			99%					99%				99%	

Rate 2

Trial Error	1	2	3	4	5	6	7	8	9	10	sum	hits	dev.
Mz	5	6	5	3	7	4	2	3	1	2	38	962	7.8
Ad	6	5	8	10	4	3	6	5	4	3	54	946	8.2
Dw	5	8	7	5	2	6	3	5	5	6	52	948	6.2
Wn	4	2	5	4	5	2	4	4	4	5	39	961	6.8
Ant	7	4	4	8	6	5	3	6	5	5	53	947	7.2
Sc	3	4	5	6	3	4	5	2	3	4	39	961	6.8
Total	30	29	34	36	27	24	23	25	22	25	275	5725	43.0
Av.	5	4.8	5.7	6	4.5	4	3.8	4.2	3.7	4.2		954.2	
Ad.	1	1.5	1.2	2	1.5	1	1.2	1.2	1.1	1.1		7.2	
%acc				94.8					96.0			95.4	

exact and final adjustments. A movement, therefore, which depends for its accuracy largely or entirely on its initial adjustment should, according to this view, rapidly become inaccurate with increased speed IF the "initial adjustment" must act in "opposition to the momentum of uniformity." The decreased interval does no good, and the speed does harm. Such a movement Woodworth illustrates by his "three target test" (op. cit. p. 49) in which the subject thrust with a pencil

alternately at three dots arranged in the form of an equilateral triangle. To quote him: "Here each initial adjustment is a repetition not of the last, but of the third before. Between, two movements of quite different direction have been made. And anything like automaticity is excluded by the angular character of the whole movement. The natural tendency to cut corners is not a help but a hindrance to accuracy. Thus the 'momentum of uniformity,' if operative at all, is so only to a slight extent. The result is that, in proportion as speed interferes with 'current control,' the accuracy is lost. There is no flattening out of the curve at the higher rates, but the error increases by great jumps, making the last part of the curve the steepest." This discussion, especially as it relates to the two factors in voluntary movement, and to the rapid loss in accuracy at the fast rates, is directly applicable to the results from the three hole apparatus.

In Table XVIII are the results of 6,000 thrusts at Rates 1 and 2, in which "secondary movements" were ruled out by instructing each subject not to correct a miss, but to go directly to the next hole, the rhythm being kept the while. By this change in procedure, the accuracy at both rates is reduced, though Rate 1 is still more accurate than Rate 2, and Rate 2 more accurate than Rate 3. The practice curves for both rates, (Diagram 12) show the same flattened appearance from trial 6 on, previously noted in the curves for the other rates. The decrease in accuracy, though still small, is much greater than the decrease for the same rates in the previous experiments. Rate 1 shows a loss of 1% in accuracy, with or without the practice effect, and Rate 2, a loss of 4.6% without, and 4% with correction for practice.

The question of the optimum rate for thrusting remains to be considered. Taking the experiment as a whole, it is clear that no matter whether the subject is left free to correct his inaccurate thrusts, or is instructed not to correct them, Rate 1, as far as errors go, gives the greatest accuracy. Rate 2 is next, and the other rates follow in order, so that we can generalize by saying that speed and accuracy are inversely related in a regular manner. We could hardly stop here, however, for the relation of increase in speed to decrease in accuracy, as well as the practice effect, have not been taken into consideration. With practice eliminated, the accuracy curve (Diagram 11) shows Rate 2 as equal to Rate 1, while the relatively slight

decrease in accuracy at Rate 3 (5.9%) and even the decrease at Rate 4 (17.3%) would certainly give these rates a claim in many sorts of everyday use of a co-ordinated movement like the one here illustrated. In some kinds of work, an employer might very probably prefer an operator of a machine who could make a movement in ½" with 6% errors in 6,000 trials to one who required twice as long to make the same movement, even if the latter were always accurate. With practice eliminated, a consideration of the time factor, therefore, would certainly shift the optimum rate from Rate 1 to Rate 2, and in some cases to Rate 3. Rate 5 is too inaccurate to be considered, as the chances that a subject will "hit" rather than "miss" are only 1 in 2—a subject hits about as often as he misses. This reduction of the accuracy to a "chance" basis obviates the necessity of attempting a faster rate.

Reference to Table XVII, shows that the six subjects tend to "bunch" more closely at Rates 1 and 5, and to "spread out," become more variable, at the intermediate rates. This is most probably due to the fact that they were equally accurate at Rate 1, and equally inaccurate at Rate 5, the middle rates tending to bring out individual differences. It is interesting to note that the A. D. does not vary much with successive trials—though practice reduces the scores of all the subjects, they hold very closely to their relative positions.

Individual differences in the accuracy of thrusting appear at Rate 1, and may be found throughout the experiment. If we rank the subjects in order for accuracy at each rate, we get the following table.

	R–1	R–1′	R–2	R–2′	R–3	R–4	R–5	Sum	Ave.	Rel. Pos.
Mz.	–	3	4.5	1	3	3	5	19.5	3.25	3
Ad.	–	6	4.5	6	5	5	6	32.5	5.42	6
Dw.	–	4	3	4	4	6	3	24.0	4.00	4
Wn.	–	1	1	2.5	1	1	1	7.5	1.25	1
Ant.	–	5	6	5	6	4	4	30.0	5.00	5
Sc.	–	2	2	2.5	2	2	2	12.5	2.19	2

Note: The (′) at Rates 1 and 2 refers to the second series.

Wn and Sc are the most accurate throughout, and Ant and Ad the least accurate. There is very little shifting of relative position, a subject evidently "hitting his stride" early, and holding it. Table XVII shows that the subjects who are least

accurate, are usually inaccurate in the first trials at the different rates; these trials have too many errors in comparison with successive trials. Observation of the thrusting of the different men suggests that some men have a "knack" for simple co-ordinated movements which others do not possess, either through lack of fine muscular control, nervousness, susceptibility to becoming easily confused, lack of patience, etc. In a long series of thrusts, with the rhythm fixed, there is little chance for luck to play much of a role, and, therefore, the subject's record is a pretty good indication of his ability in the function tested.

CONCLUSIONS

1.—Speed and accuracy in a comparatively simple co-ordinated movement, such as the one here used, are inversely related—as the speed increases the accuracy decreases.

2.—In comparison with the increase in speed the decrease in accuracy is slight, so that if time is considered as a factor, the optimum rate for both speed and accuracy, with practice eliminated, would lie at Rate 2, and in some cases at Rate 3.

3.—Practice was markedly operative in the experiment during the first 3,000 thrusts, but in the last 3,000 thrusts, if present at all it was very slight.

4.—The conclusions stated so far hold, whether the subjects are instructed to rely solely on accuracy of aim or are permitted to use secondary correcting movements.

5.—In this experiment, those men who rank high at any given rate, tend to hold their relative positions at all other rates.

2. Maze Tracing Experiment

RESULTS

In the tracing movement as required in this experiment, there was no distinction, such as that made in the previous experiment, between "primary" and "secondary" movements. No matter what the rate, there was always plenty of time for initial adjustment—the subject was given time enough to place his pen on the cross at the beginning of the pathway, and catch the rhythm before beginning the actual tracing.

This experiment requires a precise co-ordination of the muscles of the arm, hand, and fingers with the visual impression; accuracy depending, more specifically, on the skill with which the pen point can be moved along between the two guide lines. There is very little difference between the mechanics of the tracing and the thrusting movements, as here used; both are comparatively short and intermittent, and both require the same muscular and visual co-ordination. In tracing, however, the movement is shorter and made with less effort—it is also more continuous in the same general direction, less angular, than the thrusting movement.

Table XIX gives the results of the experiment in detailed form, and is arranged after the plan already described in the Thrusting Experiment. One difficulty not met in the former experiment arises when we come to estimate the relative accuracy at the different rates. In making thrusts at a bull's eye, each thrust is either a hit or a miss, and accuracy is easily figured in % of hits in a total number of thrusts. In tracing, on the contrary, the number of touches in a given movement is a variable quantity, and may range between fairly wide limits. Accuracy, therefore, can only be measured in terms of errors divided by total number of movements. This is what has been done, though it is a question whether we can assume that one touch is equal to another as a measure of failure to co-ordinate. Any weighting system, however, involves too many sources of error, variable factors, etc., and

TABLE XIX

Records in terms of touches: 6 Subjects: 5 Rates: 1020 separate movements at each rate. Accuracy is reckoned 100%—(errors÷total number of movements). Total number of movements is divided into 2 groups; in second group pratice is eliminated.

Rate 1 Met. 60

Trials	1	2	3	4	5	6	7	8	9	10	sum	dev.
Errors—i. e. touches												
Mz	3	3	0	0	1	2	2	0	1	0	12	6
Ad	1	3	2	7	3	4	6	3	2	2	33	15
Dw	1	0	1	0	2	1	0	0	0	0	5	13
Wn	4	0	0	0	0	0	0	0	0	0	4	14
Ant	6	5	1	0	4	5	5	5	7	6	44	26
Sc	2	1	1	1	0	3	2	0	0	0	10	8
Total	17	12	5	8	10	15	15	8	10	8	108	82
Av.	2.8	2	.8	1.3	1.7	2.5	2.5	1.3	1.7	1.3	18	
A.D.	1.5	1.0	.3	1.0	.8	1.1	1.2	.9	1.0	.9	13.7	
Errors			52					56				
%Acc								98.17			98.24	

Rate 2 Met. 96

Trials	1	2	3	4	5	6	7	8	9	10	sum	dev.
Errors—touches												
Mz	8	5	2	2	1	0	1	2	1	0	22	26.8
Ad	10	9	10	11	11	9	8	9	4	0	81	32.2
Dw	6	5	7	0	2	1	0	2	0	1	24	24.8
Wn	9	4	3	1	1	2	0	2	2	1	24	23.8
Ant	14	10	11	8	4	9	7	6	12	9	90	41.2
Sc	10	10	7	5	2	7	6	2	1	1	51	2.2
Total	57	43	40	27	21	28	22	23	20	12	293	153.0
Av.	9.5	7.2	6.7	4.5	3.5	4.7	3.7	3.8	3.3	2	48.8	
A.D.	1.7	2.5	2.8	2.7	2.7	2.9	2.1	2.4	2.5	1.7	25.5	
Errors			188					105				
%Acc								96.85			95.21	

Rate 3 Met. 120

Errors—touches												
Trials	1	2	3	4	5	6	7	8	9	10	sum	dev.
Mz	10	9	11	6	5	3	2	7	8	5	66	19.5
Ad	5	15	9	11	10	10	9	8	9	6	92	6.5
Dw	16	14	14	14	16	9	8	4	6	8	108	22.5
Wn	13	8	10	6	8	6	6	13	10	7	87	1.5
Ant	11	19	14	14	10	9	6	7	5	7	102	16.5
Sc	18	9	5	6	2	7	4	3	3	1	58	27.5
Total	73	74	63	57	51	44	35	42	41	33	513	94.0
Av.	12.2	12.3	10.5	9.5	8.5	7.3	5.8	7.0	6.8	5.5	85.5	
A.D.	3.5	3.7	2.5	3.5	3.5	2.0	1.6	2.3	2.2	1.7	15.7	
Errors			318					195				
%Acc								93.63			91.62	

Rate 4 Met. 160

Trials	1	2	3	4	5	6	7	8	9	10	sum	dev.
Errors—touches												
Mz.	23	20	10	12	10	13	8	7	8	7	118	46.8
Ad	21	23	19	14	6	11	11	21	13	19	158	6.8
Dw	22	14	20	18	18	18	15	21	18	16	180	15.2
Wn	20	25	12	13	8	7	10	9	10	7	121	43.8
Ant	43	41	31	25	23	26	25	23	11	18	226	101.2
Sc	20	15	20	17	7	12	17	16	11	11	146	18.8
Total	149	138	112	99	72	87	86	97	71	78	989	232.6
Av.	24.8	23	18.7	16.5	12	14.5	14.3	16.2	12	14.3	164.8	
A.D.	6.0	7.2	5.2	3.5	5.7	5.0	4.7	5.5	2.5	4.7	38.8	
Errors			570					419				
%Acc								86.3			83.84	

Rate 5 Met. 240

Trials	1	2	3	4	5	6	7	8	9	10	sum	dev.
Errors—touches												
Mz.	52	46	51	48	46	53	49	45	41	37	468	41.2
Ad	64	58	57	56	59	51	51	62	54	68	575	65.8
Dw	50	49	35	34	45	36	36	42	53	40	420	89.2
Wn	50	46	53	52	49	42	37	42	38	42	451	58.2
Ant	70	68	49	39	65	59	62	67	55	56	590	80.8
Sc	66	63	63	43	55	40	64	65	52	40	551	41.8
Total	352	330	308	272	319	221	299	323	293	278	3055	377.0
Av.	58.7	55	51.5	45.3	53.2	47	49.8	54	49	46.3	509.2	
A.D.	8.0	80	6.3	6.7	6.5	7.5	9.2	10.8	6.2	7.3	62.8	
Errors			1581					1474				
%Acc								51.8			50.1	

consequently it was deemed best to treat all errors as equal.

It is not necessary to repeat what was said in a previous connection concerning the measurement of changes in accuracy by number of errors made at different rates—nor to state again that although the tracing at Rate 1 has 1/5 as many touches as at Rate 3, it is not 5 times as accurate, but rather 1.09 times. Diagram 13 shows the corresponding changes in the two variables with increase in rate. As the speed of tracing increases 60%, the accuracy decreases 3.7% (measured from Rate 1), and when the speed increases 100%, the accuracy decreases 1.1%; an increase of 167% in speed results in a loss in accuracy of 17.6%, and at Rate 5, with an increase in speed of 300%, there is a "drop" in accuracy of 59%. In this experiment the decrease in accuracy seems to be slightly more rapid than in the thrusting experiment, though the accuracy curves for both experiments are very close together.

Practice was clearly evident at all rates, and largely concentrated in the early trials. Diagram 14 shows the changes from trial to trial for each rate, and with the exception of Rate 1 and Rate 5, practice seems to have spent itself by the time the 5th or 6th trial is reached. In the case of Rate 1, accuracy remains at nearly the same level throughout, except for a slight downward trend at trial 3; while in Rate 5, practice making for accuracy, seems about balanced by speed or interference making for inaccuracy. The dotted line (Diagram 14) drawn in on the curve for Rate 5 represents what the level of accuracy for this rate would probably be, were the trials continued long enough for these factors to equiliberate each other.

Interference has been mentioned as a factor which along with speed made for inaccuracy. It will be remembered that, in order to avoid a concentration of the practice effect at one rate, successive sheets were traced at different rates following a regular order. This plan necessitated a shift from rate to rate, and caused a majority of touches to appear in the first section of the maze, due to faulty adjustment to the changed rhythm. This interference grew gradually less as the experiment continued, and as the subject became habituated to the different rhythms. In spite of this fact, however, if we make due allowance for the interference factor, it is very true that the procedure used in the present experi-

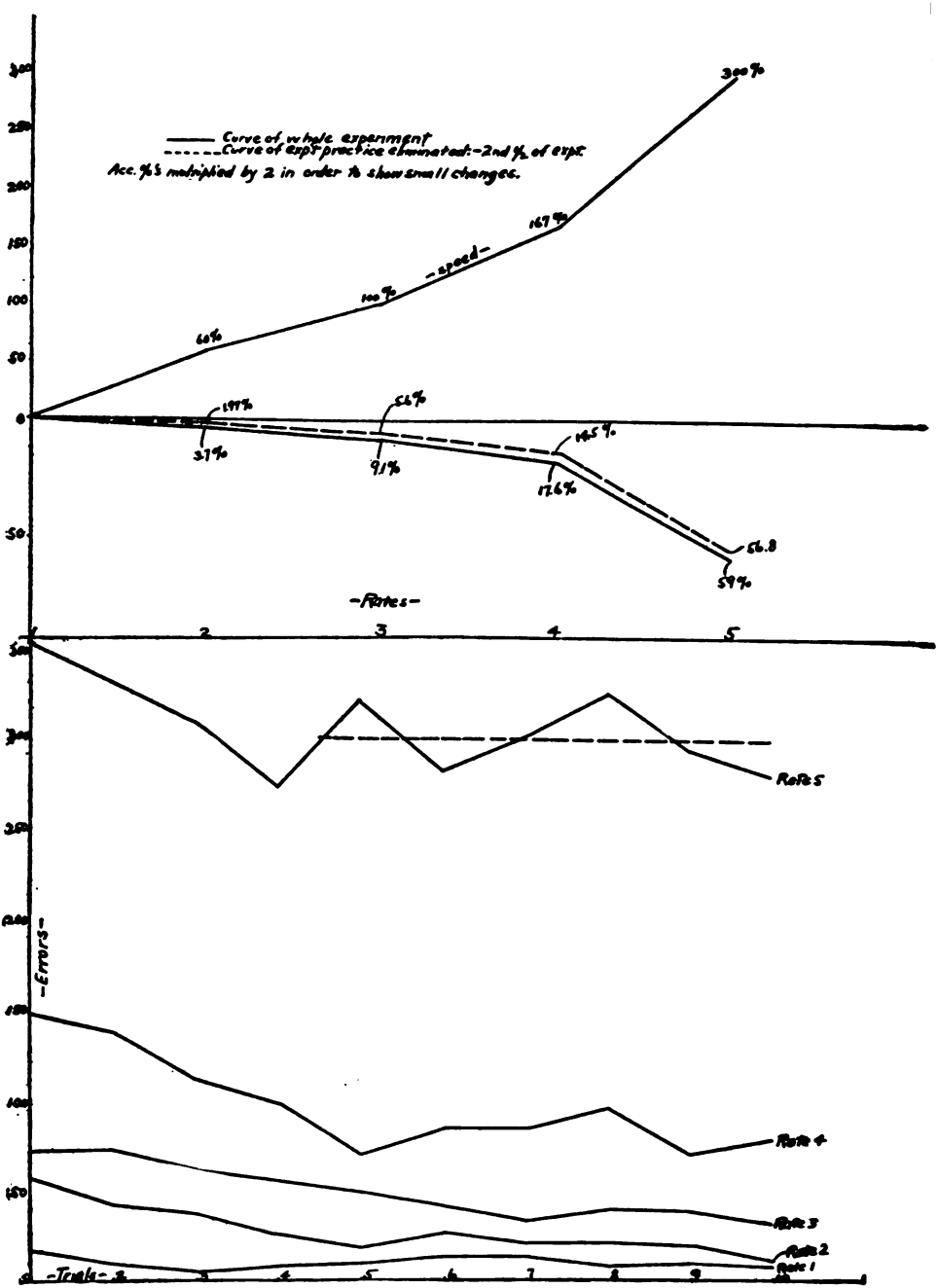


Diagram 13.—Showing the changes in the accuracy of tracing corresponding to the changes in the rate. Both speed and accuracy measured from Rate 1.

Diagram 14.—Practice curves for Rates 1, 2, 3, 4, 5,—showing changes for each trial from 1-10. Errors measured in "touches."

ment tends to make the accuracy records for any given rate lower than they would be had all the sheets at one rate been traced successively. This could not be done in this experiment, since, with any other procedure than the one used, the different rates would not have been comparable—and results would have been vitiated by a massing of practice, the avoidance of which was absolutely necessary for a fair comparison.

By grouping the trials from 6 to 10 together, I have calculated what the loss in accuracy at each rate would be, with practice eliminated. These average are represented in Diagram 13 by the dotted line, and show, except at Rate 1, an increase of between 1½%-3% in accuracy over the corresponding records for the whole experiment.

The question of an optimum rate for tracing brings up the same problem raised by the thrusting experiment. If accuracy alone is considered Rate 1 is clearly optimal. But if Speed and Accuracy both are taken into account, Rate 2, Rate 3, or even Rate 4, have claims over Rate 1. Of course Rate 5, the errors at which are more than 50% of the total number of movements, cannot be considered. Expressed in terms of increase in accuracy, Rate 1 is only 1.02 times as accurate as Rate 2; 1.057 times as accurate as Rate 3; and 1.17 times as accurate as Rate 4.

The variability of the six subjects as measured by the AD is greatest at Rate 5. Here there is an average of 509.2 touches with an AD from this average of 62.8. After Rate 5, the subjects tend to "spread out" most at Rate 4 and Rate 2 in order given, "bunching" together more closely at Rates 1 and 3. The explanation of the variability suggested in the thrusting experiment was that subjects tend to be more nearly alike at the very slow rates or the very fast rates, due to the ease of the former and the difficulty of the latter; while the intermediate rates really test the subject's ability in co-ordination, and hence are more "spread out." This same explanation cannot be carried over to the tracing experiment, without considerable modification. From Rate 1 to Rate 2, there is an increase in the variability but there is a decrease at Rate 3, due largely to Dw's large, and Sc's small increase in error. At Rate 4 there is an increased "spread" which continues over into Rate 5. The wide range at Rate 5, in contrast with the decrease in variability at the same rate in the thrusting experiment, is very probably due to the fact that,

the movement being shorter in tracing, Rate 5 is not as fast in tracing as in thrusting. A higher rate of speed might reduce the range among individuals. Practice, too, as a result of the thrusting experiment (tracing came second) might tend to carry over the individual differences from the former experiment.

The statement has been made that the functions involved in tracing and thrusting are very similar, and the inference was made that, for this reason, we should expect the same man's records in the two experiments to correspond closely. In the following table the relative positions of each subject in the two tests have been compared for each rate. (For the thrusting experiments at Rates 1 and 2, the second series have been used.)

	Rate 1		Rate 2		Rate 3		Rate 4		Rate 5		Final-Rel. Pos.	
	Th.	Tr.	Th.	Tr.	Th.	Tr.	Th.	Tr.	Th.	Tr.	Th.	Tr.
MZ.	3	4	1	1	3	2	3	1	5	3	3	2
AD.	6	5	6	5	5	4	5	4	6	5	6	5
DW.	4	2	4	2	4	6	6	5	3	1	4	4
WN.	1	1	2.5	3	1	3	1	2	1	2	1	1
ANT.	5	6	5	6	6	5	4	6	4	6	5	6
SC.	2	3	2.5	4	2	1	2	3	2	4	2	3

The correlation between the two tests is very high. Wn. in first in both tracing and thrusting; Mz, is second in tracing and third in thrusting; Sc. is third in tracing and second in thrusting; Dw. is in fourth place in both experiments; while Ad. and Ant. occupy the 6th and 5th places in thrusting, and the 5th and 6th places in tracing, respectively.

These results would surely seem to justify the inference that the functions involved in the two sorts of movement are much the same. Those factors which make for accuracy in the one experiment, determine to a large extent accuracy in the other also.

CONCLUSION

1.—In general, Speed and Accuracy are inversely related in maze tracing though the loss in accuracy is not comparable to the increase in speed.

2.—Practice, except at Rate 5, is largely concentrated in the first half of the trials at each rate. In the latter half, practice effect is very slight, or altogether negligible.

3.—When the speed of the movement is of importance, the optmum rate will probably shift from Rate 1 to Rate 2 or 3.

4.—Individual Subjects hold their positions (relative) from one rate to another with a high degree of consistency. Correlation between skill in thrusting and skill in tracing is very high, indicating clearly that the same essential factors are involved in the two functions.

5.—Some indviduals have better muscular control, are less easily confused, have greater perseverance, and more patience than others. Individual differences are largely due to these factors.

SECTION IV

RECAPITULATION AND SUMMARY

The primary object of this paper, as stated in the Introduction, was to answer certain questions having to do with the relation of speed to accuracy. These questions have all been answered, as fully as the experimental results seemed to justify under the separate experiments, and only a brief summary of the facts already given is here attempted. Each question is considered separately, in the same order in which it comes in the Introduction.

1.—Will accuracy fall off regularly as the speed increases, or is there an "optimal" point at which the accuracy is greater than at lower or higher rates? The experiments on lifted weights, linear magnitudes, and handwriting specimens have shown that accuracy is not highest when the speed or the rate is slowest, but that accuracy tends to increase gradually from the slowest rate used, 4″, to a 2″ rate, after which it falls off regularly, the loss becoming more rapid as the speed increases to 1″, and ½″. This indicates an optimal interval close to 2″ for all three experiments, a period at which the factors involved in judgment are at a point of maximal clearness. A lengthening or a shortening of this optimal time, through the introduction of irrelevant factors, or the failure of the factors involved in judgment to mature, results in either case in a loss of accuracy. In simple motor co-ordination, the speed-accuracy relation is inverse, thought not proportional. Accuracy decreases as the speed of the movement increases, though with practice the thrusting movement was made just as accurately in 5/8″, metronome at 96, as in 1″, metronome at 60.

2.—How do increases in speed affect accuracy? The loss in accuracy with increased speed is relatively very slight. In the experiment on lifted weights, the accuracy for all six observers ranged from 76.3% at a 4″ lifting rate, to 80.1% at 2″, and fell to 70.9% at the 1″rate. Thus a decrease of 75% in time interval, or an increase of 300% in the speed of lift, caused a change in accuracy of only 10%. In comparing lines, the accuracy at the 4″ exposure interval was 66.73%; this increased to 70.28% at 2″ interval, and dropped to

62.64% at ½″ interval. Again, an increase of 700% in the rate with which the cards were exposed, resulted in a change of about 8% in the accuracy. In comparing specimens of handwriting, a change of 300% in speed was paralleled by a total change of 6.3% in accuracy. With the co-ordination experiments, the changes in accuracy were much greater. As the speed of thrusting increased from 1″ to ¼″, 300%, the accuracy fell from 100% of hits to 53.5% of hits; and in maze tracing, with the same increase in rate the accuracy fell from 97.8% to 40.1%. In both of these experiments the accuracy fell off rather slowly at first increasing by leaps as the rate became more rapid.

3.—Do speed and accuracy have essentially the same relation in the perception of differences as in simple co-ordinated movements, or does the accuracy behave differently in the two cases? At first glance it would seem from the above discussion that accuracy does not vary alike in the two cases. It is doubtful, however, if we can fairly compare the speed-accuracy relationship in situations which are as different as those in the movement and perception experiments. The actual increase in accuracy as the time for perception of differences is shortened, up to a certain point, and the subsequent loss in accuracy after this point has been passed, has been explained as due to the fact that an optimal time or period—and no more—is needed for the factors involved in the judgment, visual, kinaesthetic etc.— to become maximally clear; or for the memory or set, and in the case of lifted weights, motor "Einstellung," to "carry over" from one impression to the next. In the motor co-ordination experiments, on the contrary, no comparison, memory, or judgment were required. The subject simply made the movement at designated time intervals, which, after practice became fairly automatic, although the angular character of the thrusting and the difference in direction of the tracing prevented a "momentum of uniformity"[89] from being set up. Accuracy is, therefore, largely dependent upon dexterity, skillful co-ordination of arm, hand, and eye. At the slow rates the errors are due mostly to carelessness; as the rate increases the accuracy decreases rapidly until as the "physiological limit" is reached, the time becomes so short that the neouro-muscular mechanism cannot function with any precision in the time allowed.

If we take the two groups of experiments, the perception—of difference group as rate increase from the 2″ point of maximal accuracy, and the two movement experiments as rate increases the 1″ point of 100% accuracy, we find that in both cases there is an inverse relation of speed and accuracy, though the decrease in accuracy is much greater in the movement experiments. This relatively large loss in accuracy is partly due to the fact that the times for the movements were actually much faster than the times allowed for perceiving differences. For example, the slowest thrusting or tracing time was 1″, 4 times as fast as the slowest rate in the perception experiments, and the ratio holds for the other rates. Accuracy behaves alike in the two cases, in that after a certain point, any increase in speed causes a "drop" in accuracy; but before this point is reached, increase in rates affect the observer differently, for reasons already indicated.

4.—Does the same individual hold his relative position in the different situations,—is he fast and accurate, slow and accurate, etc., or can individuals be so classified? In attempting to get at the relative standing of an individual in the 5 experiments, it has been realized that although a large number of trials were made at each rate in every experiment, that there were only six observers, and that, therefore, conclusions as to relative position are conditioned by the size of this group. In the following table the "best rate,"—the rate at which accuracy was highest,—is given for each observer, the % of accuracy, and the individual's standing in the group, considering the whole experiment. The two movement experiments have been considered together, as they gave results so closely correlated, that it did not seem worth while to consider them separately.

At first glance the differences between the records of the observer,—the range from highest to lowest,—are so small, that with one or two exceptions, we might well say that one observer was about as good as another in perceiving small differences, whether of weights, lines, or specimens of handwriting. Especially in the line experiment, the differences are so small,—only 2.9% between the highest and lowest record, that it is impossible to say any observer was definitely "best" at any rate. A further examination of the table, however,

[1]Woodworth: Accuracy of Voluntary Movement, Psy. Rev. Mon. Sup. *13*, p. 27-62.

TABLE XX

Weight Lifting

Obs.	"Best Rate"	Rel. Pos. in Exp.	%Accuracy.
Mz	3	4	81.0
Ad	1	6	74.7
Dw	3	3	81.7
Wn	3	1	84.0
Ant	1	5	80.3
Sc	2	2	82.7
		Lines	
Obs.			
Mz	3	1	73.7
Ad	3	4.5	70.8
Dw	1	3	72.1
Wn	2	2	72.9
Ant	2	4.5	70.8
Sc	5	6	70.4
		Handwriting	
Obs.			
Mz	3	1	77.81
Ad	2	4	72.95
Dw	3	3	73.14
Wn	3	5	69.90
Ant	2	2	74.38
Sc	–	–	–
		Coordination Exper.	
Obs.			
Mz	1	4.0	
Ad	1	5.5	
Dw	1	3	
Wn	1	1	
Ant	1	5.5	
Sc	1	2.0	

reveals certain definite tendencies. Mz, for example, ranks first in comparing lines and in perceiving differences in handwriting, but he drops to 4th place in the perception of differences in weight and in co-ordination. It might be expected, that skill in weight lifting would be correlated with skill in thrusting and tracing, since the former partly, and the latter almost wholly depend on muscular precision, and kinaesthetic factors. Further evidence in favor of this inference is seen in Wn's and Sc's records. These observers rank first and second respectively in the accuracy of weight comparison, and in co-ordination, while Ad and Ant rank last and next to the last. On the other hand, Wn is in 5th place in comparing handwriting specimens, and Ant is in 2nd place. Also, in estimating linear differences, Wn is 2nd and Sc is 6th, though these differences in records are largely to be discounted because of the slight range in accuracy in this experiment. Ad ranks uniformly low,—5th or 6th place—in all the experiments.

It seems that the conclusion is justified, therefore, that an observer holds his relative position in those experiments in which the functions involved are closely related: an observer who is relatively accurate in perceiving differences in weights, will also be accurate in co-ordinating simple hand and arm movements,—and the reverse is also true. There seems to be little relation between ability to perceive weight differences, or differences between linear magnitudes, or specimens of handwriting. Observers accurate in the one experiment may or may not be accurate in the other two.

A word should be said in regard to the rates used. They were entirely the result of experiment and were not predetermined. No slower rate than 4″ was deemed necessary because accuracy always increased from this rate, while Rate 5, 1″, except in the Line Experiment showed an accuracy loss sufficiently large to make further reductions in exposure time,—in order to show the general trend—unnecessary.

AN EXPERIMENTAL STUDY OF HUNGER IN ITS RELATION TO ACTIVITY

By

TOMI WADA, Ph.D.

ARCHIVES OF PSYCHOLOGY
R. S. WOODWORTH, Editor

No. 57

NEW YORK
June, 1922

CONTENTS.

Chapter		Page
I.	Introduction and General Description of the Present Study.	5
	1. Introduction	5
	2. Main Purposes	7
	3. The Apparatus	8
	4. The Subjects	10
	5. The Main Experiments	11
II.	Nature of Hunger	12
	1. Hunger Mechanism	12
	2. Activities of the Stomach	13
III.	Correlation of the Hunger Rhythm with Other Physiological Conditions	23
	1. Respiration	23
	2. Vasomotor Volume	23
	3. Salivary Secretion	23
IV.	Correlation of the Hunger Rhythm with the Sensation of Hunger	24
V.	Correlation of the Hunger Rhythm with Bodily Activities	27
	1. Bodily Movements of Men During Sleep	27
	2. Correlation of the Hunger Rhythm with Bodily Movements of Men During Sleep	27
	3. Correlation of the Hunger Rhythm with Bodily Movements of Men During Waking State	30
	4. Bodily Movements of Infants	31
	5. Bodily Movements of Albino Rats	32
VI.	Correlation of the Hunger Rhythm with Dreaming	33
VII.	Correlation of the Hunger Rhythm with Motor Activity	35
VIII.	Correlation of the Hunger Rhythm with Mental Activity	45
IX.	The Effects of Various Stimuli upon the Hunger Rhythm	54
	1. The Effect of Mechanical Stimulation	54
	2. The Effect of Drugs	56
	3. The Effect of Conscious Effort, Thought, Sight, Smell and Taste of Food	58
	4. The Effect of Nausea from Rotation	59
	5. The Effect of Electric Shocks	60
	6. The Effect of Prolonged Work	60
	7. The Effect of Reading Exciting Stories	61
X.	Conclusions	62

PREFACE

The present study of hunger and activity was undertaken at the psychological laboratory of Johns Hopkins Hospital, first in the summer of 1920, and more systematically, from February, 1921, to February, 1922. Owing to the fact that it is a new field of work for the psychological laboratory, and especially, that it was a new field for the experimenter, numerous trials and errors were made. Though further researches are eagerly hoped for and needed along this line, it is encouraging that this one and a half year's experimentation yielded certain definite results, which are presented in the following pages.

ACKNOWLEDGMENTS

Acknowledgments are due to Professor Edward L. Thorndike for guidance given to the writer in the study of general psychology as preparation for the present experiment, and for his introduction of the writer to the psychological laboratory of Johns Hopkins Hospital. The author's gratitude is sincerely expressed to Dr. Arthur I. Gates, whose encouragement and advice were immeasurable. Special acknowledgment is due to Dr. Curt P. Richter of Johns Hopkins Hospital for his valuable suggestions and assistance. To the parents of the two babies the author feels great gratitude and admiration for their sacrificial interest in the scientific work and the privileges given to the writer. To Professor R. S. Woodworth for the present publication, and to Miss Caroline Stackpole, Miss Marian Benedict, and Miss Mary Evenden, who corrected the English, and to other people who assisted the present work by becoming subjects or by reading manuscript and proof, acknowledgment is gratefully made.

TOMI WADA.

June, 1922.

Chapter I.

1. INTRODUCTION.

The need to understand the fundamental drives of human behavior is tremendously increasing in these years of social unrest, economic readjustment, and educational reformation.

Pointing out the lack of understanding as to the fundamental urges of labor problems, the late Mr. Carleton Parker[1] says that the laborers' philosophy is a stomach philosophy, and their political industrial revolt is a hunger riot. Indeed, hunger is one of the strongest of human urges. "On the same plane with pain and the dominant emotions of fear and rage as agencies which determine the action of the organism, is the sensation of hunger," says Cannon[2]. "It is a sensation so peremptory, so disagreeable, so tormenting, that men have committed crimes in order to assuage it. It has led to cannibalism, even among the civilized. It has resulted in suicide. That dull ache, or gnawing pain, referred to the lower midchest region and the epigastrium, may take imperious control of human actions."[2] Hunger not only compels the striped muscles to seek food, but urges man and animals to fight, and even to risk life itself. Hunger is such a strong drive that the term itself stands as a representative of all wishes and desires. When the effort to satisfy hunger is thwarted, the whole organism reacts to the situation, or the thwarting agent, with such hyper-tension of all organs and muscles and fibres that the excitement may lead to various types of defensive behavior. Moreover, as McDougall[3] says that all instinctive impulses may, with opposition or obstruction, give place to, or are complicated by, the pugnacious, or combative, impulse directed against the source of the obstruction. Thus hunger may become a dynamic force of destructive or constructive emotions and behaviors.

In the primitive type of man, hunger plays such a rôle that the greater part of his daily activity is directed by it. Food was the first form of property. The value of things was measured in terms of food. "Primitive migration was always after food, food for men and food for cattle, which in turn was food for men."[4] Ceremonials, religion, manners and laws grew in close relation with food. Food was the first thing desired and fought for.

[1] Parker, Carleton; Casual Laborer and Other Essays, 1920.
[2] Cannon, W. B.; Bodily Changes in Pain, Hunger, Fear and Rage. 1920. P. 232.
[3] McDougall, Wm.; Introduction to Social Psychology. 1908. P. 72.
[4] Hall, G. S.; Adolescence.

The earliest education and habit formation of children begin also with regard to food. Psychological tendencies, such as acquisition, pugnacity, rivalry, jealousy, sympathy and gratitude display themselves in connection with food, and more or less "condition" or determine the future character of the child. Thus food is a matter of great concern in infancy and childhood. Except when in need of food, a healthy infant is asleep. With the onset of hunger the sleeping baby awakes, wriggles and cries. If food is not given him he struggles more and more. His eyes wide open, his mouth moving or chewing anything at hand, he cries with all his might, hands clinched tight, face red, hot, and often perspiring, legs kicking, or drawn up tightly to his body. Often, too, his back is bent, and stretched. The whole organism is struggling. G. Stanley Hall[5] says, "the true beginning of a psychology essentially genetic is hunger, the first sentient expression of the will to live." This must be the "Elan Vital."

As soon as food is given, the infant sucks it with such power that often he cannot wait to breathe. The whole organism is in a state of readiness and tension. The swallowing reflex takes place as soon as any particle is placed on the tongue. Very soon food and sucking are coupled together, and this becomes an established habit. But before that, all kinds of efforts and errors are made to satisfy this great drive of hunger. We often see babies chewing the edge of a blanket, and more often, their own fingers.

In animal psychology it is a well known fact that rats, chickens, cats, and dogs do not take any interest in solving puzzle cage problems unless they are hungry. The varying food habits and hunger behaviors of different species of animals are very interesting in relation to their environment. As Rogers[6] did in 1916, Pavlov[7] recently proved by the following experiment that a greater part of the alimentary center is situated below the cerebro-hemisphere, and controls hunger behavior, more or less independently of the central nervous system. "A decerebrated pigeon remains immovable for hours in the middle of an abundance of food, being incapable of feeding itself. However, in such a pigeon, the activity of the alimentary center manifests itself clearly.

[5] Hall, G. S.; Adolescence.
[6] Rogers, F. T.; The Hunger Mechanism of the Pigeon and its Relation to the Central Nervous System. Am. Jour. Physiol., Vol. XLI.
[7] Pavlov, J. P.; Sur le centre de la faim. Jour. de Psychologie. April, 1921.

We nourish him, and he takes the grains in his beak. Five or six hours after the meal, the animal begins to stir, walk about, and becomes more and more active. It is easy to be convinced that it is the alimentary center that provokes that agitation. In order that he may again become immovable for a long time, it is sufficient to catch the pigeon and make him swallow a sufficient quantity of grains." Thus even when there is no cortical control, hunger leads to a greater degree of motility on the part of the animal capable of locomotion, and to a greater degree of activity of the feeding reflex. Even among the lowest types of unicellular animals and plants, that possess no nervous system, hunger manifests itself strongly enough to cause visibly increased motility and cell excitability.

2. Main Purposes.

What can we say about the influence that hunger exerts upon the activities of human beings? It is said that hunger, even during sleep, dominates and often controls the nature of dreams and if strong enough awakes the sleeper. Is this true? What is the effect of hunger upon muscular and mental work? Biologists say that fish, that are going against a river current, and butterflies, that are undergoing great physical exertion, never feed.[8] It is a common practice among students and business men to eat but little when there is important mental work to be done. Does hunger have any influence even upon the specialized functions that are measured by intelligence tests? What is the influence of emotional states upon hunger? Some animals refuse food, when in captivity, even to the point of starvation in the presence of plenty. Do men feel the same way in great emotional excitement? Such are the problems to be discussed in the following experiments.

We owe valuable guidance in studying "instinct" to William James[9]. "The older writings on instinct are ineffectual waste of words, because their authors never came down to this definite and simple point of view, but smothered everything in vague wonder at the clairvoyant and prophetic power of the animals." He says, "The strict physiological way of interpreting the facts leads to far clearer results." Such was the method employed in the present experiment.

[8] Carlson, A. J.; The Control of Hunger in Health and Disease. 1917. P. 8.
[9] James, Wm.; Principles of Psychology. 1890. Pp. 383-385.

3. The Apparatus.

1. A Stomach Tube. For the purpose of registering the activity of the stomach, a very thin rubber balloon of flexible quality, 4 cm. in diameter, and from 12 to 13 cm. in length, was chosen. This balloon was connected with a rubber tube of 3 mm. diameter and of any desired length. A very small metal tube was inserted at one end of this tube, and the rubber balloon was tied over it with silk thread. The metal tube made the connection firm, yet did not close the rubber tube. A little rubber cement was used in order to have the rubber balloon attached firmly and smoothly, so that the edges of the balloon would not irritate the throat, when the subject swallowed it. A number of such stomach tubes were kept ready. They were disinfected before use, and were kept in running water, in order to preserve the good quality of the rubber.

When the air was sucked out, the balloon and the tube were very small, and could easily be swallowed. Infants usually swallow them even more easily than adults.

2. A Recording Apparatus. This apparatus consisted of a U tube 2.5 cm. in diameter and 22 cm. in height, half filled with water. One limb of this manometer had a floating marker which consisted of a thin cork, an aluminum plate 12 cm. long and 2 mm. wide, with a thin celluloid flag on top of it, and a cork adjuster on the upper end of the U tube. The other limb of the manometer had a rubber cork from which a rubber tubing reached to a glass bottle 17 cm. high. (See Figure 1.)

When the subject was in the desired position, he swallowed the tube, which was connected with the recording apparatus. The other end of the stomach tube was connected with another rubber balloon of the same kind (B), which was enclosed in a glass bottle (D). The experimenter blew through the adjoining tube (C), which distended the balloons in the stomach and in the bottle and then closed it with a clamp. The air pressure in the bottle forced the water up in one limb of the U tube (E), and forced up a floating marker (F). The experimenter opened the clamped tube (G), and let the air escape until the water in the manometer was at the same level on both limbs, and closed it again with the clamp. Every change in pressure on the balloon in the stomach registered on the smoked paper of a kymograph. This method had certain distinct advantages over those previously used. Others connected the stomach tube directly with the

manometer. The result was that the air pressure in the balloon kept the manometer water unequal. Consequently the apparatus lost sensitivity. By the method here employed, the level of the water in each limb was the same at the beginning of the experiment and hence sensitivity was increased.

Another method employed in the present experiment was to use a double rubber stomach balloon between the layers of which a bismuth coat was inlaid. Observations through a fluoroscope made it possible to locate the exact position of the balloon in the stomach. When the balloon was inflated, the movements of the stomach could easily be observed, since the bismuth is opaque to the X-rays.

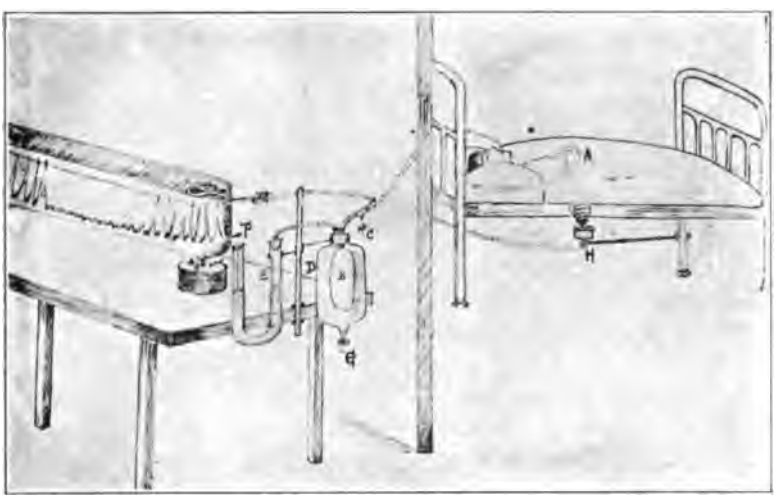


Fig. 1. The Apparatus.

(A) A stomach balloon, swallowed into the stomach. (B) A balloon connected with the stomach and enclosed in the bottle. (C) An adjoining clamped tube, to blow in the air. (D) A bottle to adjust the air pressure. (E) A manometer of U tube. (F) A floating marker. (G) A clamped tube at the bottom of the bottle. (H) A tambour registering activities on the bed.

3. In order to register the gross bodily activities of a man or an infant during the day or night, a tambour 5½ cm. (2¼ inches) in diameter, and 2¼ cm. (⅞ in.) in depth was used. This tambour was covered with a rubber dam of good quality and fairly thick. A thin metal disc, 3 cm. (1⅛ in.) in diameter was shellacked in the middle of the rubber membrane, and on this disc a 2 cm. (⅞ in.) flat-headed wooden screw, with the pointed end up, was attached by means of wax. Fairly elastic wire was

made into a loose spring about four inches in height. One end of this spring was tied with cord to the stem of the wooden screw, while the other end was attached loosely to the bottom of a spring bed. By means of a clamp on one leg of the bed, the tambour was supported, and at the same time can be adjusted according to the weight of the subject, so as to keep the tambour at the maximum of sensitivity. The rubber tube from the bottom of the tambour was connected to a Marey tambour to register the body movement on the smoked paper. In earlier experiments, from three to ten tambours were used, distributed at different points of the bed. But experience showed that one in the proper position was just as sensitive, if not more so. (Fig. 1. H.) For the experiments with infants, the tambour, without the adjusting wire spring, in direct contact with the bottom of the baby bed, or carriage, was found more sensitive.

4. The kymograph that was used was the extension kymograph, on which paper 7 feet 10.5 inches (94.5 in.) long could be placed. A speed of 20 inches per hour was usually used. Faster and slower speeds were tried, with the result that this speed was found to be the most accurate and convenient. Every 4½ hours changes of paper had to be made. When the experiment was carried on throughout the night the paper had to be changed only once, at midnight.

4. The Subjects.

T. W., a woman student in psychology.

Subject C. P. R., a research man in psychology.

Subjects H. R. B., T. W. C., and H. E. H. were graduate students in medicine in Johns Hopkins University. All of them were healthy, vigorous young men.

Patient Z., a young woman, in the Phipps psychiatry clinic, suffering from chronic hypochondria.

Patient Y., a young man of the clinic, suffering from general weakness and inability to work.

Baby Ann R., born February 13, 1921. She was nine months old, weighing 18 pounds, at the time of the experiment, which was conducted from November to January, 1922.

Baby Barbara P., born on November 10, 1921, was one month old when the experiment was begun on December 28th. She weighed 11 pounds 9 ounces at the end of the experiment, February 2, 1922.

Sixty Albino rats, of different sex and ages, were under the

writer's care from April to June, 1921. Observations on other rats were made also.

5. The Main Experiments.

The first set of experiments here presented concern themselves with the nature of hunger. Neurological study of the autonomic and the sympathetic nervous systems was carried on, as well as anatomical study of these systems, and of the stomach muscles. The dynamic phases of the problem were studied by direct measurement of the activity of the stomach. Varying forms of stomach activity were recorded on smoked paper by means of a special apparatus. The frequency and duration of contraction periods were studied. The results presented in Chapter II show the general facts concerning stomach activity.

The next step, described in Chapter III, is to find the correlation between hunger states and other physiological conditions, especially in respiration, vasomotor changes, and salivary secretion flow.

The third experiment, recorded in Chapter IV, is to determine the sensation of hunger and its relation to hunger contractions.

Chapter V introduces the measurement of the gross bodily movements of men during sleep. These results were analyzed and compared with the results of similar experiments tried during the day. The experiments were extended to infants, whose activities were measured during the day and night. Finally rats were subjected to general observation.

Mental activity during sleep was studied through the experiments with dreams which are presented in Chapter VI.

Experiments to ascertain if there is augmentation of muscular work by hunger contractions are presented next, in Chapter VII. The hand dynamometer measurements were correlated with the hunger states.

Chapter VIII deals with the correlation of hunger states with mental work in Thorndike mental examinations.

Chapter IX discusses the results obtained from a series of experiments in which various stimuli were introduced. Mechanical stimuli, drugs alleged to have power to relax smooth muscles, conscious effort, the thought, sight, and smell of food, were given. These results throw light on the origin and control of hunger contractions. Stimuli of still stronger kinds, such as electric shocks, rotation, prolonged work and reading exciting stories were introduced, and the effects of these activities upon hunger states were measured.

CHAPTER II.

THE NATURE OF HUNGER.

1. HUNGER MECHANISM.

1. A Brief Description of the Stomach.

The stomach is a pear shaped organ, situated in the epigastric region of the body, with three chief parts, the *fundus,* the *body,* and the *pyloric portion,* which is located at the right extremity of the stomach. The region where the oesophagus enters is known as the *cardiac,* and the junction with the intestine is called the *pylorus.*

The musculature of the stomach consists of non-striated fibres of three kinds, namely, the longitudinal, or outer layer, the circular layer, and the middle layer. The middle layer is lined by the submucous coat, and finally by the mucous coat.

2. Nervous Control of the Stomach.

It is generally maintained that the vagi nerves are the main afferent pathway for the hunger impulses. The vagi nerves have their nuclei in the fasciculus solitarius, in the medulla. The efferent vagi motor fibres come to the stomach from medulla nuclei. There are reasons for believing that the vagi and the sympathetic (splanchnic) nerves, which come from the coeliac ganglion, regulate the tonus of the stomach, one as accelerator, and the other as inhibitor. Some maintain that both nerves have accelerating and inhibiting fibres in themselves. Experiments have been repeatedly made to show that when either the vagi or the splanchnic nerves are cut, separately, or both together, the stomach keeps its movements without much disturbance.

There are some other nerve fibres, such as the plexus of Meissner, between the submucous layers, and the plexus of Auerbach, between the longitudinal and circular, muscular layers. These intrinsic plexus seem to be capable of maintaining the movement of the musculature, independently of the extrinsic nerve.* The real origin of the gastric hunger contractions is not yet definitely established. Some maintain the myogenic theory, and some the neurogenic theory of the intrinsic plexuses. Chemical changes in the blood also has been considered as a possible cause of the gastric contractions. Though

* Müller, L. R.; Das Vegetative Nervensystem. 1920. Pp. 127-134.

the periodicity of the hunger rhythm and the abrupt cessation of the contraction periods do not parallel with the fact that the chemical or starvation changes in the blood are more continuous, yet the hypothesis can not be altogether abandoned. Especially, such an internal secretion as that of adrenalin may act with certain periodicity upon the stomach walls and the intrinsic nerve plexuses and start the hunger contractions at the same time this chemical change is affecting the other parts of the organism.

3. Hunger Contractions Versus Digestive Contractions.

Howell* says that the chief difference in character between the digestive and the hunger contractions seems to lie in the fact that the former involves mainly the antral end, while the latter are described as starting from the region of the cardiac orifice, spreading as peristalic waves over the whole stomach.

Digestive movements occur a few minutes after the entrance of food. Small contractions start in the middle region of the stomach and run to the pylorus. The fundus end of the stomach is not actively concerned with these movements, but serves rather as a reservoir for the food, while the muscular pyloric region is the apparatus that triturates and macerates the food, and forces it out from time to time into the duodenum. Hunger contractions start in the cardiac end of the stomach, and proceed towards the fundic region. Gradually the contraction waves cover a larger and larger space. From May 10th to 14th observations were made, through a fluoroscope. of the stomach movements of the young woman patient, Z. From May 18th to 20th a young man, Subject Y., was under observation. These observations verified the above statement of the activities of the stomach, its advancing waves, and the portions involved.

2. Activities of the Stomach.

By means of the above mentioned apparatus the writer measured the activities of the stomach. When the apparatus was very delicate the following typical states were recorded:

1. The respiration pressure rhythm. When the stomach is in atonic condition there is no movement at all. Consequently the record shows only respiration. Almost uniform ups and downs *on the level* are to be seen, as shown in Figure 2. This complete atonic condition is very rare except immediately after vigorous contractions. or after a little water is taken.

* Howell; A Text Book of Physiology. 1921, P. 721.

Fig. 2. The respiration pressure rhythm.

2. The pulse pressure rhythm. When the respiration is held for a few seconds, or when the stomach pressure gets so low that the floating marker goes down to the limit and does not record respiration, the pulse rhythm is recorded. (See Fig. 3.)

Fig. 3. The pulse rate rhythm.

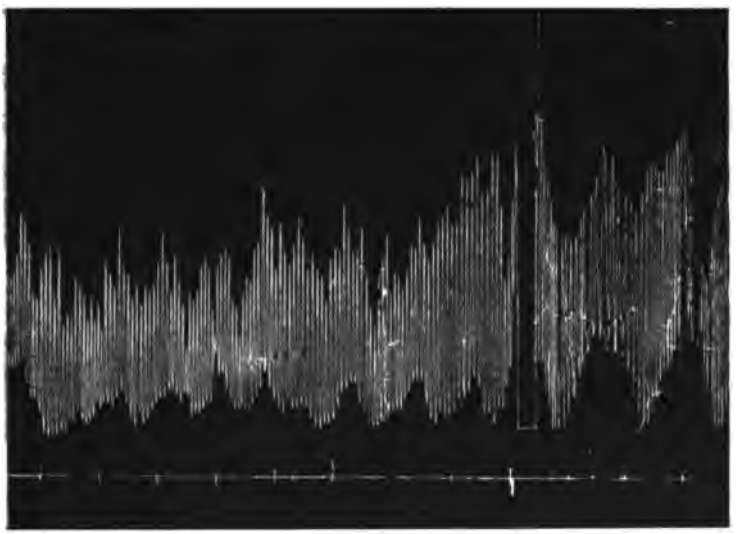
3. The tonus rhythm. The stomach is usually in tonus condition, and often with tonus waves of great uniformity. (See Fig. 4.) These tonus contractions of the fundus gradually increase in amplitude, as well as in rate. (See Fig. 5 A, B, C.)

Fig. 4. The tonus rhythm.

4. The powerful hunger contractions. Gradually increased tonus reaches to the hunger contractions. The duration of this contraction

Fig. 5. A., B. and C.

A. Beginning of tonus rhythm.



B. Gradual increase of amplitude of tonus rhythm.

C. Gradual change of rate and amplitude of tonus rhythm.

is usually from 30 to 40 seconds. (See Fig. 6.) The recorded amplitude of the contractions depends upon the size of the stomach balloon and of the manometer, as well as upon the extent to which the balloon is distended. In the present experiment, however, these factors were controlled so as to be as constant as possible. Quite often records reached to twelve centimeters in amplitude, the maximum of the apparatus. Contractions were still more powerful, and pushed the marker to the top, and water kept running over the manometer for some minutes.

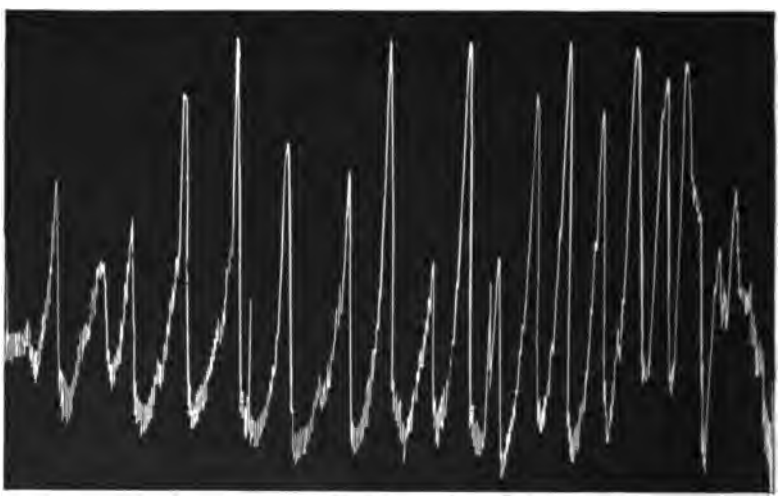


Fig. 6. The contraction rhythm.

5. The tetanus ending. The powerful hunger contractions *sometimes* end with tetanus, or prolonged contractions. (See Fig. 7.) This is especially the case with prolonged starvation. After the strong contractions, an abrupt drop into atonic, or weaker tonus conditions, often follows. (See Fig. 8.)

The strength of the hunger contractions may be measured in terms of the amount of water pushed up by the gastric pressure. A contraction that recorded a height of one inch on the smoked paper pushes 20cc. of water in a U tube of which diameter was 2.5 cm. The strongest contraction that the apparatus could record was 5 inches in amplitude, which had power to push up 100 cc. of water.

Tables I and II contain results of experiments during the night and the day. They explain the duration of quiescence and tonus, the number and duration of contraction periods, and the amplitude of the contractions. (See Figs. 9, 10.)

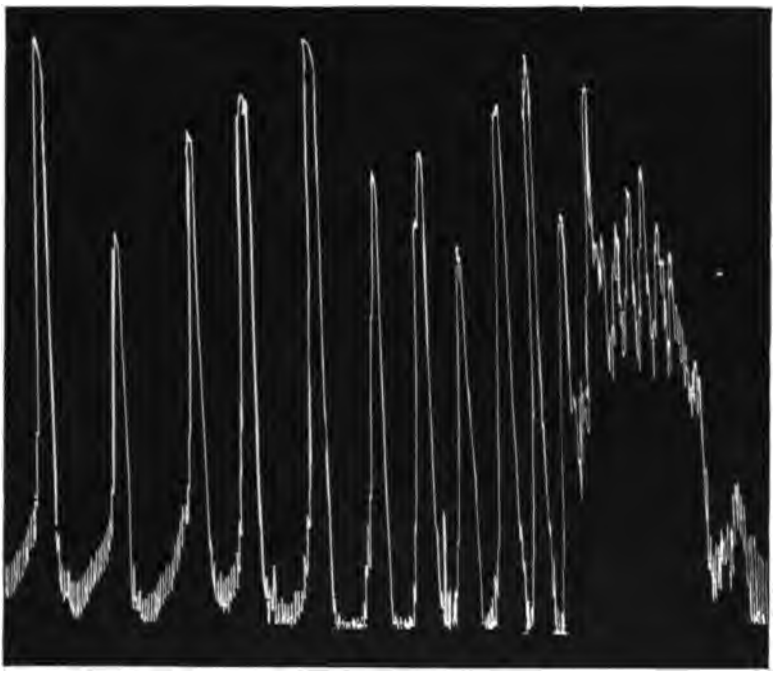


Fig. 7. The tetanus ending.

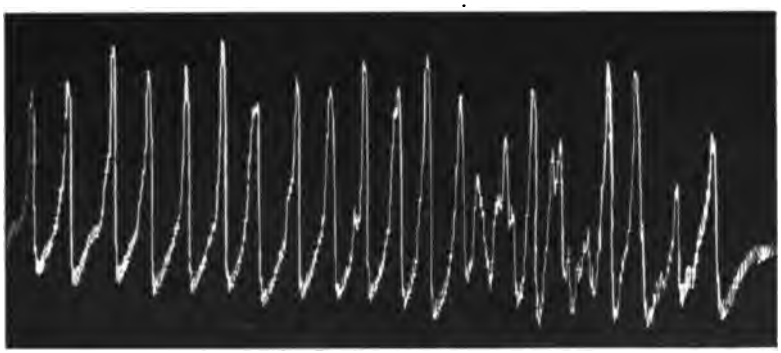


Fig. 8. The abrupt drop from contraction to quiescence.

Distribution of these different states of activity varies a great deal according to the individual's general physical condition, his rate of metabolism, caused by physical and mental work, and his food habits. However, a rough average of all records show the following features:

A. Gradual increase of the size of the contraction waves ending with tetanus.

B. Two contraction periods between which a quiescent period lies.

Fig. 9. A. and B.

from 2 to 3 hours after a meal there appear tonus waves which last from 30 minutes to one hour. They increase in amplitude and rate. and become gradually stronger contractions. An average number of the contractions is between 20 and 40 during a period which extends from 30 to 60 minutes.

Fig. 10. A. and B.

A. Recurrence of contraction periods.

B. A long contraction period.

TABLE I.

Showing the Activity of the Stomach for Various Subjects During the Night.

Date 1921	Subj.	Time Started	Duration of Quiescence	Duration of Tonus	Contraction Period	Duration of Contrac.	Aver. Amp. of Contrac.	No. of Contrac.
Mar. 23	B	10:40	:40 min.	:35 min.	12:15-12:30	:15 min.	1.5 in.	6
				:30	1:00- 1:30	:30	2.	12
			:30	:30	4:15- 5:45	:30	3.25	15
			:45	:15	5:30- 6:15	:45	2.	6
Mar. 24	B	10:30	2:15	2:00	3:45- 4:45	1:00	4.25	40
					5:15- 5:30	:15	2.	8
Mar. 25	B	11:30		:20	11:50-12:10	:20	1.25	12
				:15	12:25-12:45	:20	1.50	10
				:05	12:55- 1:25	:30	1.50	32
			:45		2:05- 2:55	:50	2.5	49
			:15		3:20- 3:55	:35	1.5	20
			:30		4:00- 4:40	:40	3.	31
Apr. 13	B	10:30	:15	2:45	1:30- 2:00	:30	1.5	10
				1:45	3:15- 3:30	:15	1.75	10
					4:15- 4:45	:30	3.75	23
					6:30- 6:45	:15	2.75	10
			:15	:45				
Apr. 20	B	10:30	:30	1:30				
				:30	2:00- 2:30	:30	2.5	19
			1:30		5:10- 5:15	:15	2.	15
					5.30- 5:45	:15	3.75	22
				:30	6:30- 7:00	:30	4.	18
					7:00- 7:30	:30	4.5	19
Mar. 25	H	10:45			1:20- 1:35	:15	2.	7
			:15		1:45- 2:20	:35	2.	15
			:10		3:45- 4:15	:30	4.25	20
					4:30- 5:00	:30	3.	9
Mar. 28	H	11:00						
Mar. 30	H	10:30	:30	1:15	12:15-12:30	:15	2.	10
			:45		1:15- 1:30	:15	2.	12
				:15	1:45- 2:15	:30	3.	14
			:20		2:30- 2:45	:15	3.25	11
			:30		4:10- 4:30	:40	3.5	18
			:15		4:45- 5:15	:30	3.5	8
				:15	5:15- 5:45	1:00	4.5	21
			:15					
June 10	H	11:50	:30		12:30- 1:15	:45	2.	30
				:30	1:35- 1:40	:05	2.	8
			:20		2:00- 2:30	:30	2.	15
					2:30- 3:15	:45	4.5	45
			:45	:25	4:25- 5:30	1:05	5.	47
			:45		6:20- 6:50	:30	5.	25
June 11	H	11:19	:30	:30	1:00- 3:20	2:20	4.5	145
			:30	:10	4:00- 4:30	:30	3.5	26
				:30	5:10- 5:25	:15	4.	18
			:45		6:10- 7:10	1:00	4.5	45
June 7	C	11:10		2:00	1:00- 1:30	:20	2.5	20
			:10		1:40- 2:00	:20	2.25	16
			:30	:30	4:30- 4:50	:20	4.5	16
			:30		5:30- 6:00	:30	3.5	19
			1:00	:15	7:40- 8:00	:20	2.5	16
June 8	C	11:50		1:15				
			:30		1:30- 2:30	1:00	3.	40
			1:30		4:00- 5:00	1:00	2.5	40
			:30	:10	5:30- 6:10	:40	3.5	34
			:10	:20	6:40- 7:00	:20	3.	7
				:10	7:05- 7:20	:15	2.	8
June 9	C	11:50	:30		1:00- 1:25	:25	2.	41
			:45	:20	2:20- 2:30	:10	2.	11
				:15	2:45- 2:55	:10	2.5	9
				:10	3:05- 3:10	:05	2.	7
				:20	3:30- 4:15	:45	2.	30
					4:20- 5:00	:40	5.	43
			:30	:15	5:45- 6:10	:25	3.	28
			:10	:20	6:30- 7:10		3.5	32
			:35					

TABLE II.

Showing the Activity of the Stomach for Various Subjects During the Day.

Date	Subj.	Time Started	Duration of Quiescence	Duration of Tonus	Contraction Period	Duration of Contrac.	Aver. Amp. of Contrac.	No. of Contrac.
1920								
July 30	W	3:35			3:40- 4:15	:35 min.	3. in.	12
			:05		4:30- 5:00	:30	3.	20
			:30		5:45- 6:25	:40	3.	21
			:30	:45	8:00- 8:30	:30	3.	19
				:20	8:50- 9:40	:50	3.	32
1921								
Apr. 10	B	10:30	:10	2:00				
			:10	1:00	1:30- 1:45	:15	2.5	9
			1:00	1:00	4:20- 4:30	:10	2.	8
				:30	5:00- 5:50	:50	3.5	31
			1:00					
June 6	B	9:45	1:00	1:00	11:30-11:40	:10	3.	3
Apr. 9	H	3:40	:10	:10	4:10- 4:30	:20	1.0	15
			:10	:10	4:50- 5:15	:25	1.5	22
			:15	:20	5:50- 6:20	:30	1.5	25
Apr. 14	H	4:40	:30	:30	5:40- 5:45	:05	2.5	2
Apr. 18	H	4:30		:30	5:00- 5:30	:30	2.5	26
			:10	:20				
			1:00					
Apr. 22	H	5:00	:10		5:10- 5:50	:40	3.5	25
			:10					
			1:20					
Apr. 24	H	9:00	1:00	:20	9:00- 9:30	:30	4.5	16
					11:20-11:35	:15	4.2	14
			:45		12:35-12:50	:15	4.2	12
			:40	:30	3:00- 3:15	:15	4.2	16
			:30	:15				
			2:00	1:00	5:30- 6:00	:30	4.	17
			1:00					
			:30					
Apr. 26	H	4:40	:30	:30	5:00- 5:15	:15	2.7	15
			1:30					
May 8	H	9:00	:30	:20				
			:15	1:25	12:00-12:30	:30	2.	15
				:15	12:45- 1:25	:45	4.5	28
			:35	:30	2:30- 2:45	:15	2.	12
				:45	3:30- 3:40	:10	1.5	6
			:50	:15	4:45- 5:00	:15	2.5	17
			1:00					
May 21	H	4:00		:20	4:20- 4:50	:30	2.	28
				:20	5:10- 5:40	:30	2.5	25
				:15	5:55- 6:10	:15	2.	12
June 9	H	9:00	:30	2:30	12:00-12:40	:40	3.	27
			:20		1:20- 1:55	:35	4.5	22
			1:00	1:35	4:30- 5:00	:30	3.5	21
			:15		5:15- 6:00	:45	4.5	34
Apr. 3	C	11:30	2:20		1:20- 1:30	:10	2.	6
			1:00	:50	3:30- 3:50	:20	2.2	11
				1:10	5:00- 6:00	1:00	3.5	52
Apr. 15	C	4:15	1:15	:45			2.2	3
Apr. 16	C	3:30	1:00	1:30				
Apr. 17	C	10:15	:30	:10	11:10-11:40	:30	3.	20
			1:30	:30	1:10- 1:25	:15	3.	14
Apr. 19	C	4:20	1:00	:30				

TABLE II. (*Continued*)

Date 1921	Subj.	Time Started	Duration of Quiescence	Duration of Tonus	Contraction Period	Duration of Contrac.	Aver. Amp. of Contrac.	No. of Contrac.
Apr. 21	C	4:55	:30		5:30- 5:40	:10	4.	10
			:30	:30	5:10- 5:55	:45	4.2	21
Apr. 23	C	1:30	:30	:30	1:30- 2:10	:40	3.5	46
					2:45- 3:20		3.	42
			:20	:20	4:00- 6:10	2:10	3.7	22
Apr. 25	C	4:33	:45		5:30- 5:40	:10	2.	7
				:20	5:45- 6:00	:15	2.5	13
				:45				
Apr. 27	C	6:52	:30					
			:30					
May 15	C	10:05	1:00	1:00				
			1:00	:20	1:20- 1:40	:20	2.	12
			1:00		2:40- 3:00	:20	3.	16
				:20	4:30- 5:00	:30	2.	3
June 5	C	1:45	:40	1:30	5:15- 5:45	:30	3.	26
June 7	C	9:30	1:35	:15	11:20-11:45	:25	2.	18
			:25	:10				
			:10					

Chapter III.

CORRELATION OF THE HUNGER RHYTHM WITH OTHER PHYSIOLOGICAL CONDITIONS.

1. Respiration.

The stomach balloon registers very sensitively the effect of respiration upon the stomach. Another method of measuring respiration on the chest by means of Marey's pneumograph was tried on Subject C from 1:45 to 6 P. M. on June 4th. All results show that there was no specially deep or shallow breathing at the time of the contraction.

2. Vasomotor Volume.

For four hours a record of the glass plethysmograph on the right arm of Subject C was taken simultaneously with that of the stomach. In periods of tonus conditions of the stomach, the rhythms of the plethysmograph record correlate closely with the ups and downs of the tonus. The vasomotor changes shown in the plethysmograph records looked as if they followed closely the ups and downs of the stomach record. But the powerful hunger contraction waves did not correspond to the rhythm of vasomotor changes recorded by the plethysmograph.

3. Salivary Secretion.

In the experiment on July 30th, Subject W went through a twenty hour starvation period. On the occurrence of each contraction period there was an abundant secretion of saliva. Her mouth became so filled with the secretion that she had to expectorate several times. This never happened in quiescent periods. An improved method of registering the salivary flow of Stenson's duct was devised by Dr. C. P. Richter, together with the present writer, by joining to the metal cannula for collecting secretion, a long, narrow glass tube, which was placed on a yard stick. The measurement of air bubbles through the tube was recorded as the secretion flowed. By using this method, together with the stomach tube, one can measure the correlation between the salivary secretion and the state of hunger. Owing to the discomfort of keeping both pieces of apparatus in the mouth for a long time, this experiment was abandoned.

In the experiment of conditioning salivary secretion to the sound of a bell, by using carbolic acid, lemon juice and chocolate, it was found that immediately after a meal conditions were unfavorable for the experiment because little saliva could be obtained for several hours. Pavlov reports an experiment by Nikifrovsky, of retarded conditioned reflex of the salivary secretion of a dog. Introduction of acid solution three minutes after the appearance of a strong light was repeated, so as to establish this three minutes' retarded conditioned reflex. "It is at 5 P. M. that we distribute food to our dogs: if we make an experiment of retarded reflex at 10 A. M., salivary secretion does not appear until the end of three minutes. But if we bring over the same experiment at 3 or 4 P. M. we cannot obtain the *retardation.*" This fact shows that the gland is more ready to secrete saliva when the dog is hungry. Thus Pavlov points out the close relation between the salivary secretion center and the hunger center. It is a common observation that people have abundunt saliva in the mouth when they are hungry. The writer's experience agrees with the statement of Carlson* that there is a rhythm of the salivary flow parallel to the gastric hunger contraction rhythm. He goes further to say that each hunger contraction is accompanied by a brief gush of saliva from the duct.

Chapter IV.

CORRELATION OF THE HUNGER RHYTHM WITH THE SENSATION OF HUNGER.

The apparatus was so arranged that the subject pressed an electric magnetic button, which registered the appearance of the sensation of hunger. At the same time, the activity of the stomach was recorded. This is the well known Cannon-Washburn experiment, and was repeated in the present study only to confirm the result.

On April 23rd, Subject C came to the laboratory at 1:15 P. M., without luncheon, and his sensation of hunger was recorded in the above mentioned way. The experiment was carried on up to 6 P. M. He pressed the button at three different contraction periods, as follows: 10 signals at the 3:20 contraction period, 5 at 4:10, and 6 at 4:50. All of the signals occurred at the height of a contraction. No sensation of hunger was recorded at quiescent periods.

* Carlson, A. J.; "Control of Hunger in Health and Disease". 1916. Pp. 90-91.

Fig. 11. The sensation of hunger (the bottom mark) recorded simultaneously with the hunger states.

Subject H underwent the same experiment, on April 24th, for 9½ hours, starting at 9 A. M. His record was as follows: 6 signals at 11:30, a tonus period, 2 at 12, at the beginning of a contraction period, 4 signals at 1, at the end of the contraction period, 5 at 1:25, 4 at 1:35, both in tonus periods, and 4 signals at 5:30, at the height of contractions.

These results show that the sensation of hunger occurred simultaneously with the contraction of the stomach. However, observations made in these experiments showed that, when the contraction became chronic from lack of food for several hours, the sensation became diffused, and the subject failed to appreciate each hunger contraction. In the first case of the above mentioned experiments, Subject C had one long continuous contraction period, from 4 to 6 P. M. He signaled only 11 times at the heights of the largest contractions, while the number of contractions counted over 140.

This experiment was repeated on May 16th with a young woman patient, Z.,* whose complaint was that she was never hungry because her stomach was weak and could not digest any food. She lay down on a bed, and pressed a button in the same way as the other subjects. The experiment was started at 10 A. M. She had a contraction period from 10:20 to 11:30, and another one at 12:35, which continued until 2:10, lasting one hour and thirty-five minutes. These contractions were of great amplitude and recurred rapidly. In this case of incessant and chronic contractions, also, the patient failed to perceive some of the weaker contractions.

Patient Y., a young man, was the subject of this experiment on May 17th and 20th. He had morning and noon hunger contraction periods. He discriminated each sensation of hunger at the height of each contraction.

A sensation of dull weakness, headache, and sometimes of nausea, is felt when the hunger pang is left unsatisfied for a long time. This seems to vary greatly in different individuals. Within the knowledge of the writer, some individuals say that they never had a hunger pang in their life, while, on the other hand, others speak of having had, at times, so strong a hunger pang in the epigastric region that they could not sit still, but had to go out to get something to eat.

The origin of the sensation of hunger was thus established as being the muscular contractions of the stomach.

* This experiment and some others following were made possible by the co-operation with Dr. C. P. Richter.

Chapter V.

CORRELATION OF THE HUNGER RHYTHM WITH BODILY ACTIVITIES.

1. Bodily Movements of Men During Sleep.

From March 9th on, every night for a week, Subject R slept on the bed that registered all his bodily movements during sleep. The stomach tube was not employed in this study. The tambour system was made so sensitive that even his finger movements were registered at the beginning of the experiments. The results were studied by putting on a chart the number of body movements made. It was found that the activities fell into distinct groups, and that these groups came with certain periodicity. Figure 12 shows a part of the periodical groups of the activities. Another interesting discovery was that, when the subject had taken a glass of milk before going to bed, he was much less active in the early part of the night than when he went to bed feeling hungry, but without the milk. Compare Fig. B, which is the record of the night when he had the milk, with C and D in Fig. 12. These facts suggested the possible correlation between stomach activity and general bodily activity.

2. Correlation of Hunger Rhythm with Bodily Movements of Men During Sleep.

A second group of experiments was started on March 23rd when the subject had become thoroughly accustomed to swallowing the stomach tube. Subject B swallowed the tube and then went to bed in one of the quiet rooms of the laboratory, where the ventilation and heating were favorable. The tubings from the bottom of the bed, together with that from the stomach, were led through a hole in a thick door into the next room, where the experimenter adjusted them to the recording apparatus. The machine operated throughout the night without disturbing the sleep of the subject. Fig. 13 shows the results. The subject had dinner at 6 P. M., studied until 10:30 and then came to the laboratory to have the experiment started at 10:45. The stomach was in a quiescent state for one-half hour, then it gradually began to have more and more tonus rhythm. At 12:15 there appeared some bodily movements, none of which had appeared hitherto. At 12:30 the experimenter observed that the body move-

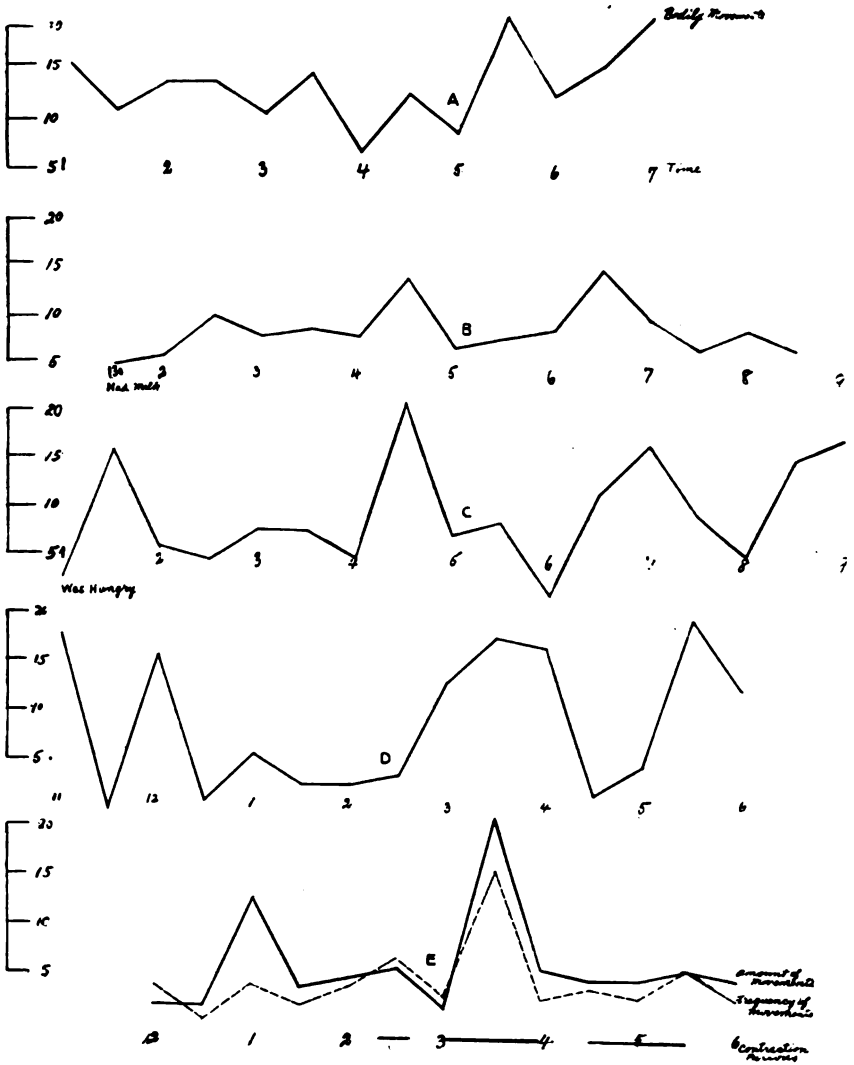


Fig. 12. The gross bodily movements in quantitative measurement.

ments occurred simultaneously with the stomach contractions. A few minutes later the stomach fell again into the quiescent period, only to resume a tonus condition within fifteen minutes. The second contraction period began five minutes before 1 A. M. Most of the body movements occurred at the heights of stomach contractions.

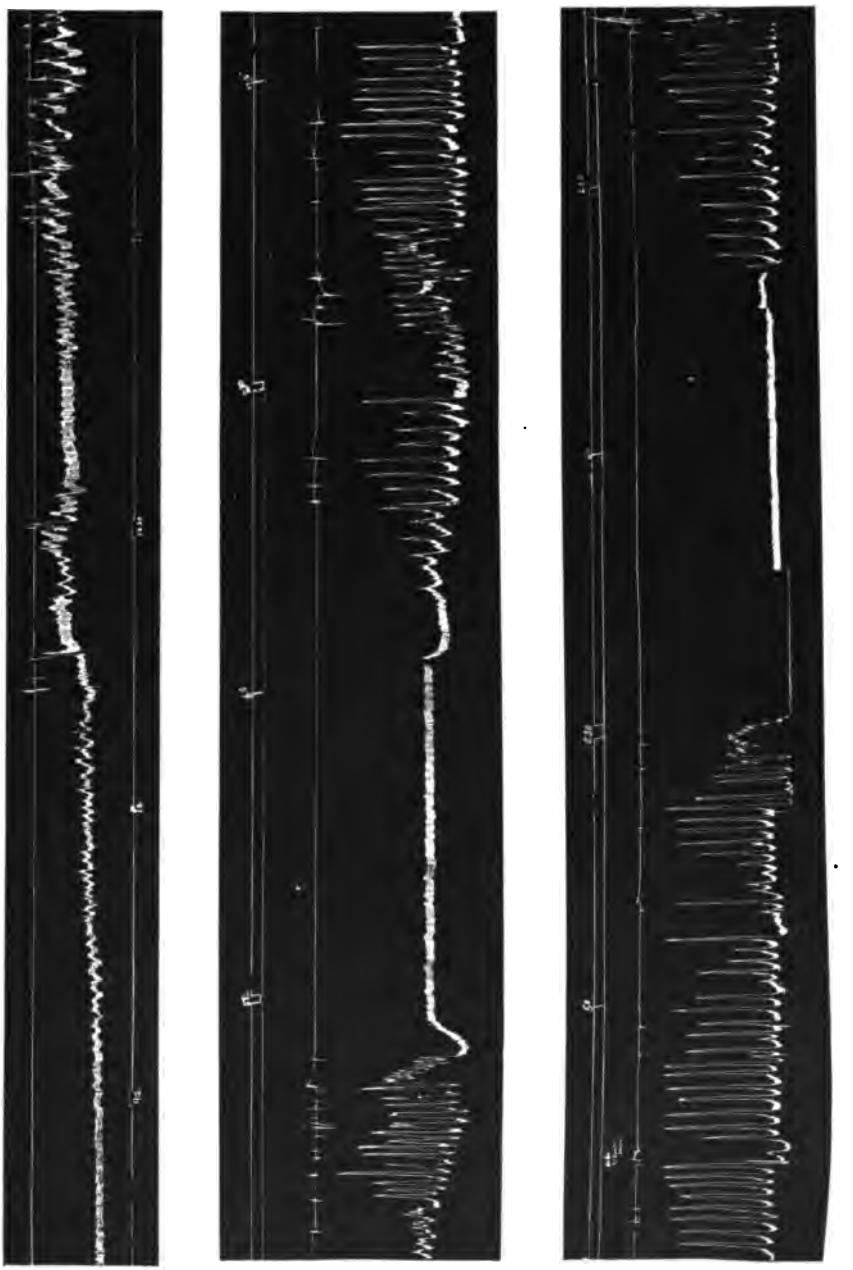


Fig. 13. A, B, and C. Bodily movements recorded simultaneously with hunger states.

There were two more definite contraction periods at 4:15-45, and 5:30-6:30. At the end of the last contraction the subject got up of his own accord, and the experiment was stopped. The subject said that he had slept well, without discomfort from the swallowed tube. The experiment was repeated on the following five nights. With slight modifications, bodily activities of men during sleep were recorded on thirteen nights. Photographs of actual records testify to the simultaneous occurrence of the bodily activities and hunger contractions. (See Fig. 13.)

3. Correlation of the Hunger Rhythm with Bodily Movements of Men During Waking State.

On April 23rd a record of the bodily movements of Subject C during the day was taken. He had eaten nothing since breakfast. At 1:15 P. M. the experiment was started, the subject lying on a couch, and reading "Moon Calf." He had contraction periods from 1:30 to 2:10, and from 2:45 to 3:20, also from 4 to 6:10. Bodily activities occurred almost simultaneously with the stomach contractions, as in the night experiments. (See Fig. 14.)

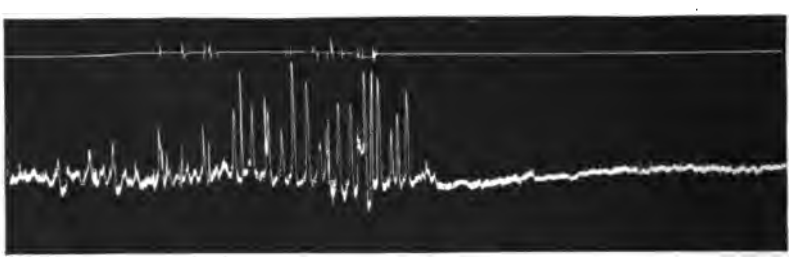


Fig. 14. Bodily movements of men during waking state.

On May 8th, Subject H had no breakfast, and was experimented on from 9 A. M. to 6 P. M. in the same way. He had the following contraction periods: 12-12:30, 12:45-1:25, 2:30-2:45, 3:30-3:40, 4:45-5:00 P. M. There were numerous bodily movements during the contraction periods, while practically none of them appeared at quiescent times. Since this was a whole day experiment, with five contraction periods, the result was confirmatory in showing rhythmical activity periods during the day, when the subject was relatively free from external stimuli.

4. Bodily Movements of Infants.

The tambour apparatus was so set underneath a small bed as to register the bodily activity of Baby Ann. From November 9th, the night activities were taken. It was necessary to have a baby whose feeding habits were regular. Baby Ann had been brought up under ideal conditions. Since she was ten months old her feeding periods had remained unchanged. They were as follows: 7:30 to 8 A. M., 11:30 to 12 M., 3:30 to 4 P. M., and 8:30 to 9 P. M. She received 8 ounces at the first feeding, then 10 ounces, again 10 ounces and at the last feeding 8 ounces of milk. She was put to bed at 9 P. M. in a quiet nursery, and was not touched until the morning feeding period. Records show very marked periodicity of activity; about once in forty-five minutes she wriggled in her sleep, and towards morning activity increased, but without changing the periodicity. These experiments were repeated for ten nights and the results confirmed the periodicity. (Fig. 15.)

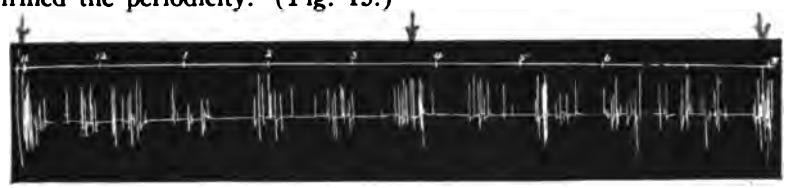


Fig. 15. Activities of infants. Baby Ann (9 months old), during the night. 11 P. M.—8 A. M.

Baby Barbara slept in a carriage instead of a crib. The tambour, which was placed underneath the body of the carriage, registered the baby's activity sensitively. Since she was very young there was no marked difference between the activities of day and night. She slept on, excepting when she was fed at 6:30 and 10:30 A. M., and at 2:30, 6:30 and 10:30 P. M. For some time after the night feeding the baby was, according to the record, very quiet. At intervals of about once every 40 minutes, she made bodily movements which increased in frequency and vigor toward the morning until the next feeding, which came at 6:30.

These periods depend upon the quantity and quality of food given, as well as upon the health conditions and the digestive capacity of the child. However, we may well predict regular activity periods as the time for feeding draws near. In the case of Baby Ann, change of diet by a small amount of oatmeal and orange juice in the course of the ten day experiment did not affect the activity and hunger periods. As the age increases food habits change, and consequently, the bodily activity periods.

Day activity records were taken for three consecutive days. The infant Ann was kept in the nursery with several toys, while any outside stimulus, such as a sight of the parents, was avoided. The feeding periods were strictly kept. The results showed that the day activities also fell in periods of about 45 minutes.

5. Bodily Movements of Albino Rats.

About 200 Albino rats were kept in the laboratory for observation and experimentation. Some 60 white rats were under the writer's care and observation from April to June, 1921. Thirty of them were 3 weeks old, and 15 were 3 months old, while the rest were 5 months of age in April. About ⅓ of each group were females. They were kept in cages and fed once a day at 12 M. Their hunger behaviors were closely observed. At the time of feeding, the rats appeared to be extremely hungry. The young ones, especially, pushed each other at the gate, climbed on top of each other, and jumped up and down on the wire cloth of the cage. The older ones sniffed and smelled the milk, and bit the wire at the gate with their teeth. After the trial in the maze, they were provided with food, which made them more excited. They fought, they screamed, and finally they snatched away a piece of bread and ate it most eagerly. After they ate enough, some of them would take pieces away from others and pile them up in a corner, often under the saw dust. The older ones would go to the nook and sleep, while the young ones played and ran after each other for a while and then would take a nap. Even the older ones woke up several hours after the feeding, and the activities of the rats gradually increased toward the next feeding period. By the measurement of his rats, Dr. Richter[1] found that the free activities of the rats came with certain periodicity. To the bottom of a triangular cage were affixed three tambours which registered on the smoked paper by means of tubing all the activities of a rat. When the rat was kept in a quiet and dark room, spontaneous activities fell in definite groups. Until about 7 days after birth, records showed no definite periodicity.[2] Activities were quite constant. But after a rat had been fed sufficiently to keep it from getting hungry for a time, it would not feed for about 30 minutes. Thus, on the 7th day, the record began to show periodicity. There was a definitely active period about once every 30 minutes. As the size of the rat, as well as the stomach increased, these periods came less frequently—once

[1] Richter, C. P.; Behavioristic Study of the Activity of the Rat. 1922.
[2] The author was permitted to mention some part of the unpublished experiment by Mr. G. H. Wang.

every 40 to 50 minutes, but the amount of activity became greater. The rat often became very vicious when hungry. Very old rats, however, were observed to be much less active, and less frequently active.

A suitable stomach balloon was made, and several trials were made to measure the stomach activity of a rat. On April 30th, the stomach tube was successfully put in a large sized and very tame rat. But the rat made every possible effort to pull the tube out. Under such abnormal conditions it was impossible to secure any good record, except a faint respiration mark.

Nevertheless the correspondence of hunger and activity periods of white rats under laboratory conditions is rendered very probable by the above results.

CHAPTER VI.

CORRELATION OF THE HUNGER RHYTHM WITH DREAMING.

The extent to which the gastric hunger contraction affects the mental status of the sleeping man was investigated in connection with dreaming.

On June 8th, Subject C came to the laboratory having had dinner at 6 o'clock. At 11:50 P. M. he went to sleep with the stomach tube. At 1:30 A. M. there was a contraction period lasting an hour. At 2:20 A. M., that is, at the end of the contraction period, the experimenter opened the door quietly and asked the subject if he had dreamed. The answer was, "I wasn't dreaming." At 4 A. M. there was another contraction period lasting one hour. At 4:25 A. M. the subject was asked the same question, concerning dreaming. "Not that I know of," was the answer. From 5:30 to 6:10 and from 6:40 to 7:00 and from 7:05 to 7:20, there were contraction periods, but no question was asked.

June 9th. Subject C. The contraction periods, as follows: From 1:00 A. M. to 1:25, 2:20 to 2:30, 3:05 to 3:10 and from 3:30 to 4:15. From 4:20 to 5:00, 5:45 to 6:10, 6:30 to 7:10. At 4:15 in the contraction period, Subject C was asked about dreams. He answered, "I had a dream, but have forgotten it." At 7:10 the answer was "Not that I know of."

June 10th, Subject H starting experiment at 11:50 P. M., had contraction periods from 12:30 A. M. to 1:15, from 1:35 to 1:40, from 2:00 to 2:30, from 2:30 to 3:15 and from 4:25 to 5:30. At 4:45 he was asked the question about dreaming. It was ten seconds

before the subject was sufficiently awake to answer. He said that he had dreamed at about 4:30, as follows. "By some reason or other there were red spots, an inch in diameter, scattered on the ground and as I walked toward them they came up a foot high and so thickly that I could not walk through them." The next morning he explained that he had been working at blood examinations for a month.

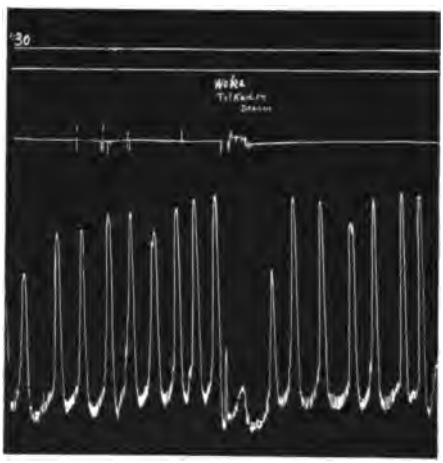


Fig. 16. Experiments on dream. Woke in the middle of the contraction and subject had a dream.

June 11th, at 3:31, during a contraction period, Subject H was questioned concerning dreams. He answered, "I had a dream, but don't remember it." The experimenter questioned him further by saying, "Do you remember what kind of dream it was? Whether pleasant or unpleasant." The subject replied, "Somewhat unpleasant; I may recall it." After a few minutes he said, "I may describe it partly: How do you want it, situation?" "Yes." "I was on a mountain, a day's journey from civilization; each day, before I went up I took my raincoat, some food, etc. After I got there, I hung those things in a place where a foreman or superintendent kept things. By some reason or other, things were disappearing there. Whenever an earthquake occurred, things were missed. One day my things only disappeared, and I was coming back with some other men, but I did not have a raincoat and got soaking wet, but tried to be cheerful saying, 'Oh, no matter, it doesn't make any difference', and I came down." Since this was a time when many students were thinking

of going home for the vacation, the experimenter asked whether it was the subject's home mountain. "No," he said, "it was not my home state mountain nor anywhere I can recall. That is all there was, thus far." The subject fell asleep and had more contractions. There was no dream in 4:00 to 4:30 contraction period, when asked at 4:22. At 5:00 the subject dreamed that when he removed the tube his throat began to bleed and much blood spurted out: this was during the contraction period. He had another contraction period from 6:10 to 7:10 A. M.

Out of seven experiments, with two subjects, on four nights, there were four dreams and two negative answers, in the contraction period. No dreams were discovered in the quiescent period.

Thus far, the observations showed that there was greater tendency toward dreaming during the contraction periods than in the quiescent. Waygandt states that starving persons dream more than usual in their sleep. (Carlson '17. P. 87.)

Chapter VII.

CORRELATION OF THE HUNGER RHYTHM WITH MOTOR ACTIVITY.

In order to measure efficiency in motor activity during the hunger contraction and quiescent periods, the hand dynamometer was chosen. Other motor tests such as tapping, steadiness, co-ordination tests, and typewriting were used with three subjects for a period of three weeks. They were found to involve many variables among which the effects of practice, fatigue, and varying degrees of interest were especially prominent. The improved Smedley hand dynamometer proved the best measurement of the maximum of grip. There was no perceptible influence of fatigue or lack of interest when experiment was undertaken at intervals of two minutes. Adaptation to the test was provided by practice for several days by Subjects B, C, H and W.

On April 18th, Subject H, without luncheon, swallowed the stomach balloon and the experiment was begun at 4:30 P. M. In order to avoid fatigue, the subject was to pull the hand dynamometer only when the order was given. The experimenter watched the record very closely and when a contraction or a quiescent period established itself definitely, she gave the order for the subject to be ready and then to pull the hand dynamometer which registered the score in terms of kilograms. Care was taken not to let the subject know the score he made at any time, but he was told to do his best. The

measurements were made, on this day, at one contraction period, which appeared between 5 and 5:30, and at a quiescent period, between 5:30 and 6. At 7 P. M., after the subject had dinner, it was resumed until 8 P. M. The scores are presented in Table III with the averages and $P.E._A$'s.

The formula is $P.E._{A.} = \frac{S.\ D.}{\sqrt{N}} .6745$

On April 22nd, Subject H's grip power was measured at one long contraction period which lasted for 50 minutes, and at quiescent periods and also after dinner for 45 minutes. On April 24th, he started this experiment at 9 A. M., without food since the previous evening. There were four contraction periods, two in the forenoon and two in the afternoon. Altogether, eleven groups of measurements were obtained, one of which was taken after dinner. Table III contains all of H's scores, the averages and $P.E._A$'s.

TABLE III.

The scores of hand dynamometer of Subject H on April 18th, 22nd and 24th. The scores measured at hunger contraction periods are indicated by stars. The calculation of averages and P.E.A's are based upon the added score of the right and the left hand.

Time	Right	Left	R. & L. Added	Average	σ	P.E.$_A$
April 18,						
4:30 P. M.	48*	44*	92*			
	45*	43*	88*			
	48*	45*	93*			
	49*	41*	90*			
	49*	45*	94*	91.40*	2.15	.65
	48.5	41	89.5			
	45	43	88			
	49	41	90	89.17	1.47	.57
		D I N N E R				
7:00 P. M.	49	44	93			
	50	41	91			
	47	40	87			
	50	42	92			
	45	42	87			
	46	43	89	91.50	2.34	.64
April 22,						
5:00 P. M.	50	41	91			
	50	38	88	89.50	1.50	.72
	50*	41*	91*			
	50*	45*	95*			
	52*	42*	94*			
	49.5*	41*	90.5*			
	48*	42*	90*			
	49*	44.5*	93.5*			
	51.5*	43*	94.5*			
	50*	41.5*	91.5*	92.5*	1.84	.44
5:30 P. M.	50*	42.5*	92.5*			
	50*	42*	92*			
	50*	44*	94*			
	50*	42*	92*			
	49*	43*	92*			
	49*	43*	92*			
	48*	41*	89*			
	51*	42*	93*			
	50*	42*	92*			
	51*	42*	93*			
	49*	42*	91*	92.05*	1.21	.25
6:00 P. M.	47	40	87			
	46	42	88			
	47	41	88			
	46	40	86			
	46	42	88	87.40	.80	.24
		D I N N E R				
6:45 P. M.	46.5	42	88.5			
	49.5	40.5	90			
	50	40	90			
	49	40.5	89.5			
	46	38.5	84.5			

TABLE III. (*Continued*)

Time	Right	Left	R & L Added	Average	σ	P.E.A.
7:00 P. M.	46	37	83			
	47	40	87			
	47	39	86			
	47	39.5	86.5			
	49	39.5	88.5			
	49	41	90			
	48	38.5	86.5			
	47	36.5	83.5			
	48	36	84			
	47.5	39.5	87			
	46	39.5	85.5			
	47.5	39	86.5	86.85	2.23	.36
	43.5	40	83.5			
	41	39.5	80.5			
	43	39.5	82.5			
	43	39	82			
	44.5	38	82.5			
	45	37	82			
	44	38	82			
	44	37	81			
	42.5	35	77.5			
	40	34	74			
	42	35	77	80.41	2.82	.57
April 24, 10:00 P. M.	51*	42*	93*			
	51*	41*	92*			
	50*	40*	90*	91.67*	1.25	.49
	49	38	87			
	49	39	88			
	48	42	90			
	49	40.5	89.5	88.62	1.19	.40
	48	37.5	85.5			
	48	37.5	85.5			
	46	39	85			
	48.5	38	86.5			
	43	36	79			
	46.5	37	83.5			
	47	35.5	82.5			
	46.5	37.5	84			
	43	37.5	80.5			
	46	39.5	85.5	83.75	2.30	.49
	47*	41.5*	88.5*			
	48*	40*	88*			
	49*	39*	88*			
	48*	39*	87*			
	48*	41*	89*			
	49*	40*	89*	88.25*	.69	.19
	46	34	80			
	46.5	39	85.5			
	45	38.5	83.5			
	47	39	86			
	48	39	87			
	47	37	84			
	46	36.5	82.5			
	44	37.5	81.5			
	44.5	37.5	82			
	45	39.5	84.5	83.65	2.07	.44

TABLE III. (*Continued*)

Time	Right	Left	R & L Added	Average	σ	P.E.A.
12:30 P. M.	48*	42*	90*			
	47*	38*	85*			
	47*	42*	89*			
	47*	39.5*	86.5*			
	45*	41*	86*			
	47*	43*	90*			
	48*	38*	86*			
	48.5*	43*	91.5*			
	47.5*	42.5*	90*	88.22*	2.21	.50
1:30 P. M.	47	41.5	88.5			
	43	41	84			
	44	41	85			
	46	38.5	84.5			
	47	41	88			
	45	41	86			
2:30 P. M.	41	39	80			
	47	41	88			
	45	41	86			
	47	43	90			
	47	40	87	86.09	2.60	.53
	50*	44*	94*			
	50*	45*	95*			
	50*	43*	93*			
	49*	42*	91*	93.25	1.48	.50
3:00 P. M.	47*	42*	89*			
	50*	45*	95*			
	51*	42*	93*			
	51.5*	42*	93.5*			
	50*	44*	94*			
	50*	45*	95*			
	50*	42.5*	92.5*	93.14*	1.90	.48
2:30 P. M.	48	41	89*			
	49	41	90			
	48	42	90			
	49	42	91			
	49	43	92			
	49	41	90			
	47	43	90			
	48	41	89			
	49	42	91			
	48	42.5	90.5	90.25	.87	.19
4:00 P. M.	48	43	91			
	47	44	91			
	50	44	94			
	45.5	44	89.5			
	50	44	94			
	46	41	87			
	49	41	90			
	50	43	93			
	44	41	85			
	46	39	85			
	45	43	88	89.77	3.10	.63

TABLE III. (*Continued*)

Time	Right	Left	R & L Added	Average	σ	P.E.$_A$
5:00 P.M.	44*	42*	86*			
	45*	43*	88*			
	49*	45*	94*			
	48*	44*	92*			
	47*	38*	85*			
	49*	39*	88*			
	48*	42*	90*			
	48*	44.5*	92.5*			
	48*	43.5*	91.5*			
	47.5*	41*	88.5*			
	48*	42*	90*	89.59*	2.66	.54
		D I N N E R				
7:00 P.M.	41	42	83			
	46	42	88			
	47	42	89			
	46	36	82			
	46	41.5	87.5			
	44.5	40	84.5			
	46.5	38	84.5			
	43	40	83			
	47	40.5	87.5			
	45	43	88			
7:30 P.M.	44	43	87	85.82	2.35	.48

On April 19th and 21st, Subject C underwent the same procedure. In all, there were obtained five groups of the contraction period scores, three quiescent and two groups of the after dinner scores. The results are to be seen in Table IV.

TABLE IV.

The scores of hand dynamometer of Subject C on April 19th and 21st. The scores measured at hunger contraction periods are indicated by stars. The calculation of averages and P.E.A's are based upon the added scores of the right and the left hands.

Time	Right	Left	R & L Added	Average	σ	P.E.A.
April 19.						
4:20 P. M.	51*	41.5	92.5			
	49	41	90			
	50	40.5	90.5	91.00	1.08	.42
	49*	43*	92*			
	49*	44*	93*			
	45*	42*	87*			
	50*	41*	91*			
	50*	41*	91*			
	49*	36*	85*			
5:00 P. M.	49*	42.5*	91.5*	90.07*	2.70	.69
	49	40	89			
	48.5	40	88.5			
	48.5	39	87.5			
	45	38	83			
	45	36	81			
5:30 P. M.	47	34	81			
	45	39	84			
	48	40	88			
	47	37.5	84.5			
	47	40.5	87.5			
	50	39	89			
	49	40.5	89.5	86.04	3.04	.59
6:00 P. M.	51*	40*	91*			
	52*	44*	96*	93.50*	2.50	.12
		D I N N E R				
7:00 P. M.	47	36	83			
	40	33	73			
	45	37	82			
	45	35	80			
	43.5	33.5	77			
	46	36	82			
	46	33	79			
	41	33.5	74.5			
	46	36	82			
	45	37	82			
8:00 P. M.	47	37	84			
	44	36.5	80.5			
	48	40	88			
	48	35	83			
	49	39.5	88.5			
	49.5	38	87.5			
8:30 P. M.	50	38.5	88.5	82.03	4.43	.72

TABLE IV. (*Continued*)

Time	Right	Left	R & L Added	Average	σ	P.E.A.
April 21,						
4:55 P. M.	39*	34*	73*			
	50*	42*	92*			
	48*	43*	91*			
	47.5*	41*	88.5*			
	52*	39*	91*			
	51*	40*	91*			
	50*	37*	87*			
	51*	42*	93*			
	50*	41*	91*	88.6*	5.77	1.30
	47	38	85			
	50	40	90			
	51.5	37	88.5	87.83	2.09	.81
	51.5*	40*	91.5*			
	51*	35*	86*			
	50*	35*	85*			
5:30 P. M.	53*	36*	89*			
	51*	35*	86*			
	50*	36*	86*			
	48*	38*	86*			
	50*	40*	90*			
	50*	38*	88*	87.50*	2.11	.47
	49	37	86			
	51	39	90			
	48	36	84			
	47	36	83	85.75	2.68	.90
	45	32	77			
	46	35	81			
	47	35	82			
	50	36.5	86.5			
	45	32	77			
	49	36	85			
6:00 P. M.	48	35	83	81.64	3.39	.86
		D I N N E R				
7:00 P. M.	43	33	76			
	49	34	83			
	48.5	34	82.5			
	50	36	86			
	49	35	84			
	50	33	83			
	52	33	85			
	49	34	83			
	47.5	37.5	85			
	45	34.5	79.5			
	47	32	79			
	47	33	80			
	47	33	80	82.00	2.78	.52
	50	33	83			
	48.5	33	81.5			
	44	30.5	74.5			
	48	33.5	81.5			
	47.5	35.5	83			
	46	36	82			
7:30 P. M.	48	34	82	81.07	2.74	.70

Figs. 17 and 18 present the curves of the hand dynamometer scores of the two subjects. The averages, which are schematically presented by the horizontal dotted lines, are higher, without exception, during the hunger contraction periods than at the preceding or following quiescent periods. Subject C's scores at the contraction periods on different days were added and the average taken is 89.98 kilograms ($P.E._{A.} = .55$). The average of all his scores at quiescent periods and at after dinner periods are, respectively, 85.64 kg. ($P.E._{A.} = .49$) and 81.84 kg. ($P.E._{A.} = .40$). C's average at the hunger contraction periods is 4.32 kg. higher than at quiescent periods, and 8.14 kg. higher than after dinner periods. Subject H's scores on three different days were calculated in the same way. The average of his scores at contraction periods is 90.91 ($P.E._{A.} = .22$). The average of his grip power at quiescent and after dinner periods are respectively 87.11 ($P.E._{A.} = .28$) and 85.42 ($P.E._{A.} = .40$). This individual can pull, therefore, at the time when the hunger contractions are present, 3.80 kilograms more than when his stomach is in non-active state, and 5.49 kilograms more than when his stomach is filled with food or is digesting it.

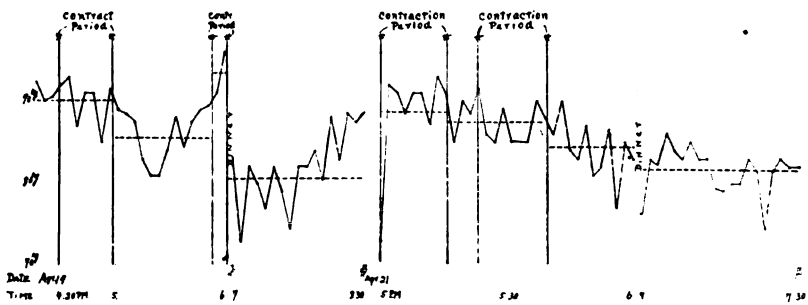


Figure 17.

The curves of the hand dynamometer scores of Subject C. A star (*) indicates the hunger contraction periods and a double cross (#) the after-dinner periods. A horizontal dotted line indicates averages.

The above experiment proves that hunger has the positive effect of augmenting the efficiency in motor activity. In other words, the rhythms of the hunger contractions and ups and downs of fluctuation in the motor efficiency, correlate with each other very highly.

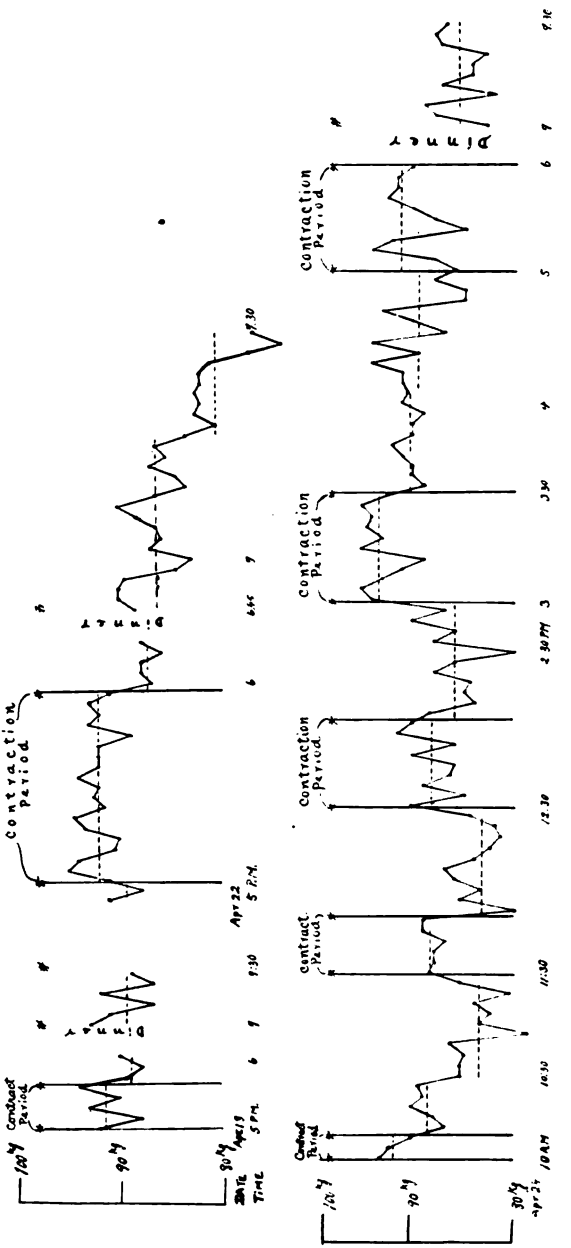


Fig. 18.

The curves of the hand dynamometer scores of Subject H. A star (*) indicates contraction periods and a double cross (#) the after-dinner periods. A horizontal dotted line indicates the averages.

CHAPTER VIII.

CORRELATION OF THE HUNGER RHYTHM WITH MENTAL ACTIVITY.

The fifteen forms of Part I of the Thorndike Intelligence Examinations were selected for the purpose of measuring the effect of hunger contractions upon mental activity. The reason for this selection was (1) that these tests have various kinds of mental functions, involving many types of materials such as words, digits, geometrical figures, etc.; (2) that these tests are supposed to be relatively free from the effect of practice; and (3) that there are fifteen different forms of tests, the relative difficulties of which have been determined.* The substitution or association tests, with different sets of numerals as keys, were first tried for several hours each, with two individuals, and it was found that the effect of practice was very great and also that equality in difficulty of the keys could not readily be obtained. Consequently, this substitution test was abandoned and the Thorndike intelligence tests were taken up.

Each form of the tests was carefully studied, and the fifteen forms were put in such an order that the same kinds of tests should not be given in succession. In giving these tests, the following two problems were kept in mind: First, whether this prolonged hard mental work would affect the vigor and frequency of the hunger contraction, and second, whether there would be any change in efficiency in doing this kind of mental work, according to the presence or absence of the hunger contractions.

On May 15th, Subject C came to the laboratory without having had anything to eat since supper the previous evening. At 10 A. M. he sat before a well-lighted desk in a quiet experimental room and his stomach activities were recorded. The experimenter watched closely and gave him an order to begin a given form and when the time was up, gave the direction, "Even if you have not finished test ——, begin test ——." At the

* Thorndike, Edward L.; Equality in Difficulty of Alternative Intelligence Examinations. Psychol. Review, 1921.

end of 30 minutes, which is the maximum time allowed for one form of tests, the subject was permitted to rest for three minutes and then he was given the next set. The forms A, B, F, C, G, K, D, M, H, L, I, O, E, and J were given up to 6 P. M. Form N was given at 8:05 P. M. after the subject had had a hearty dinner following 24 hours' starvation. In the early part of the morning, in spite of the fact that nothing had been eaten for over sixteen hours, the stomach was practically quiet. There was a one-hour quiescent period followed by a tonus period at 11 A. M. which lasted one hour and dropped into another long quiescent period. No contraction appeared until 1:20 P. M. when there were 12 contractions of the amplitude of 2 inches extending over a period of 20 minutes. One hour of quiescence followed. At 2:40 a period of 3 inch long contractions appeared and the final hunger period came at 4:30 and lasted for 30 minutes, though the amplitude of the contraction was small. After dinner the stomach was in perfect quiescence.

The accomplishment of Subject C on the tests is presented in Table V, together with the state of hunger and the expression of the subjective feeling. Fig. 19 shows the curve of his scores. The contraction periods are indicated with stars. Two of the highest scores occurred at the contraction periods. During the period when the subject reported feeling of fatigue, the stomach was quiet and the scores were relatively low, i. e. at 11 A. M. at a quiescent period, he reported, "Heaven knows I am glad to leave that," and his score was 125; at 3 P. M. in an quiescent period, he felt "tired" and said, perspiring, "I don't want to do any more," and his score was 111; at 4 P. M. he reported headache, while his stomach was in perfect quiescence and he scored 120. After the meal at 7 P. M., the subject felt much refreshed and reported, "Much easier to work now. I feel I can do ten times better." But his achievement in mental work was low, 123. Special fatigue was never reported during the contraction periods when mental efficiency was greatest. Fig. 22 shows a rapid fall in mental efficiency and the appearance of feelings of fatigue immediately following the contraction periods. The fact that the feelings of fatigue are not responsible for the effectiveness of work, is indicated by the low performance following the dinner. The evidence indicates that the hunger contraction goes hand in hand with maximal performance.

TABLE V.

Thorndike Intelligence Examination

May 15, Subject C.

Form	Time started	Hunger state	Score		Corrected score*	Time taken	Subjective feeling
A	10:05 A. M.	Quiescent	123	0	123	30 min.	
B	10:38 " "	"	123	+4	127	30	
F	11:12 " "	"	125	0	125	30	"Heaven knows I am glad to leave that."
C	11:45 " "	"	129	−1	128	29	
G	12:21 P. M.	"	129	+2.5	131.5	30	
K	12:55 " "	"	127	0	127	30	
D	1:28 " "	Contraction	142	−4	138	27′ 40″	
M	2:01 " "	Quiescent	124	+1	125	30	
H	2:40 " "	Contraction	126	−2	124	29′ 30″	
L	3:13 " "	Quiescent	110	+1	111	29′ 40″	"Tired", perspiring, "I don't want to do any more."
I	3:48 " "	"	128	+0.5	128.5	30	
O	4:21 " "	"	117	+3	120	30	"Headache."
E	4:57 " "	Contraction	137	0	137	30	
J	5:30 " "	"	131	−0.5	130.5	30	
N	8:05 " "	After Supper	122	+1	123	30	"Much easier to work now. I feel I can do ten times better."

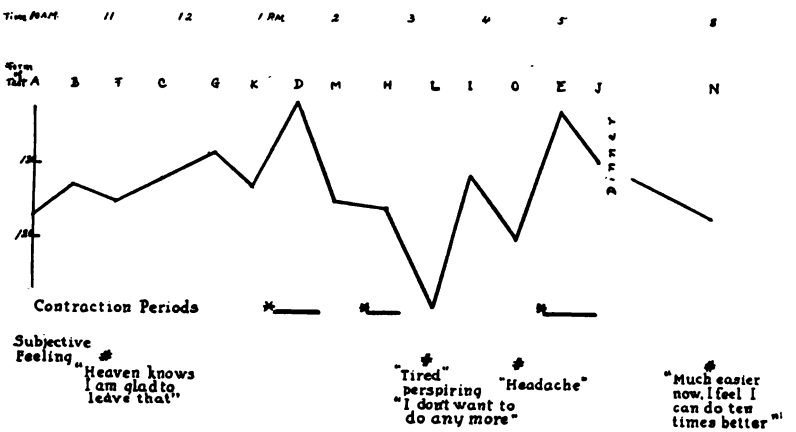


Fig. 19.

A diagram showing the curve of efficiency in the mental tests of Subject C, together with the hunger contraction periods and the subjective feeling reported. A star (*) indicates the contraction period and a double cross (‡) the subjective feeling reported.

* Thorndike, Edward L.: Equality in Difficulty of Alternative Intelligence Examinations. Psychological Review, 1921.

Subject H took the same tests on June 9th, beginning at 10:35 A. M. He had had no food since the previous evening. The successive order of the tests was changed so as to bring the earlier half of the whole series later and vice versa, in order to compare with the records for C. In the order of H, M, L, I, O, E, J, B, F, C, G, K and D, thirteen tests were given up to 6 P. M. After dinner, at 7, form N was given, while the last test was

TABLE VI.

Thorndike Intelligence Examination

June 9, Subject H.

Form	Time started	Hunger state	Score		Corrected score	Time taken
H	10:35 A. M.	Quiescent	135	−2	132	30 min.
M	11:08 " "	"	125	+1	126	30
L	11:43 " "	"	145	−1	144	30
I	12:19 P. M.	Contraction	151	+0.5	151.5	30
O	1:23 " "	"	136	+3	139	30
E	1:55 " "	"	140	0	140	30
J	2:30 " "	Quiescent	148	−0.5	147.5	28′15″
B	3:12 " "	"	144	+4	148	28
F	3:39 " "	"	146	0	146	29′25″
C	4:10 " "	Contraction	152	−1	151	25′45″
G	4:38 " "	Quiescent	126	+2.5	128.5	29′13″
K	5:08 " "	Contraction	147	0	147	28′10″
D	5:37 " "	"	148	−4	144	26′20″
N	7:00 " "	After Dinner	142	+1	141	29′35″
A	11:13 " "	Contraction	151	0	151	29

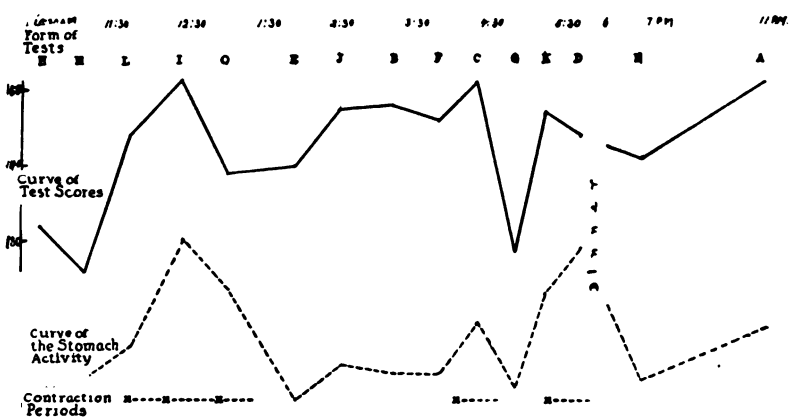


Fig. 20.

A diagram showing the correlation of the hunger rhythm with the efficiency of Subject H in the mental tests. A star (*) indicates the hunger contraction period.

tried at 11:13 before the subject retired. Table VI shows the hunger state, the scores, and the time taken. Half of the group were finished within the time-limit, the shortest of them all being 25 minutes 45 seconds. But no special credit was given for it. Expressions of the subjective feeling were not obtained in this case. Fig. 20 shows the correlation of the curves of the intelligence test scores of Subject H, and of the stomach activity which was drawn from the measurement of amplitude of each stomach activity on the graphic record. An examination of these curves reveal that four out of five contraction periods were accompanied by highest test scores, 144, 151.5, 151, and 147; while the remaining one contraction period had a test score of 139, which is a little below the average, 142.5 (P.E. = 6.43). Study of the curve shows a rapid rise in efficiency, with the onset of the contractions to a maximum at about the middle of the period. Accompanying the decrease in the intensity of the contractions to the quiescent period, is a similar decline in the efficiency of the mental work. The two curves, in fact, are nearly parallel. Following the 6 o'clock dinner, a period of quiescence was followed by a gradually increasing stomach activity which did not culminate in acute contraction when the experiment was discontinued at 11:30 P. M. The curve of mental achievement is a close parallel.

In order to test further the correlation of the subjective feeling with the actual efficiency, form D was given to Subject C on June 9th, at 11:55 P. M. when he said that he was "tired, having been busy until so late." The set E was given at 8:08 A. M. when he woke from sleep and felt "Very good." Scores were 142 and 147 respectively. Subject H repeated this experiment on June 11th and took D at 11:19 P. M., E at 7:35 A. M. He scored 145 in D and 148 in E. All of these tests were given during contraction periods. Table VII shows the details. During the "tired" periods, the scores are slightly lower, but the time spent is less, that is, there appears a tendency to hurry the work with, probably, no loss of efficiency per unit of time. This finding is in harmony with the other studies of the influence of feelings of fatigue[1-2].

[1] Thorndike, Edward L.; Educational Psychology, Vol. III.
[2] Gates, Arthur I.; Diurnal Variations in Memory and Association. Variations in Efficiency during the Day together with Practice Effects, Sex Differences and Correlations.

TABLE VII.

Thorndike Intelligence Examination

Form	Time started	Hunger state	Score		Corrected score*	Time taken	Subjective feeling
June 9, Subject C.							
D	11:55 P. M.	Contraction	146	–4	142	24'25"	"Tired being busy until so late."
E	8:08 A. M.	"	147	0	147	28'30"	"Very good."
June 11, Subject H.							
D	11:19 P. M.	Contraction	149	–4	145	23 min.	"Very tired from working hard a whole day."
E	7:35 A. M.	"	148	0	148	30	"Feel fine."

The amount of difference in achievement in the Thorndike tests between contraction and quiescent periods may be illustrated in another way. The Thorndike Intelligence Examination tests were given to the medical school class to which the two Subjects C and H belonged. Form D was given at first as a trial and immediately after it, form E was taken. Fig. 21 shows a surface of frequency based upon the measurements of this class of eighty-two graduate medical students with form E. The average of the scores of the total class is 117.5, the median deviation (P.E.) being 8.81 and S.D. 11.77.

Subject C's score in this class was 106. This was before the fifteen forms of tests were given. All of his scores in the tests taken successively later on, are plotted in a curve in Fig. 21. Subject H made scores 114 and 116 in forms D and E respectively, when he took it with his class. All of H's scores are on the diagram in Fig. 22. Fig. 23 shows diagramatically the relative position of the two subjects in their class. C's average is 126.3, while H's is 144.2. The curve for Subject C is the same as in Fig 21. H's curve was altered from the one following the successive order of tests, into one according to the order of tests that C took. This changed but little the general outline of H's curve. The variability of performance is so great, that the scores of the single individual range over half, or more than half, of the range of the whole class.

Part of the individual variability is due to practice and to the inequality of the several forms of the tests, although correc-

	For the class	For Subject C	For Subject H
Range	59 (83-142)	41 (106-147)	37.5 (114-151.6)
S. D.	11.77	8.26	11.0

tions for the latter were made. The learning curve is extremely irregular and Figs. 19 and 20 show that the variations are correlated with conditions of the stomach, rather than with the degree of practice. That the shape of the curve was not caused wholly by the inequality of the tests is indicated by the fact that the

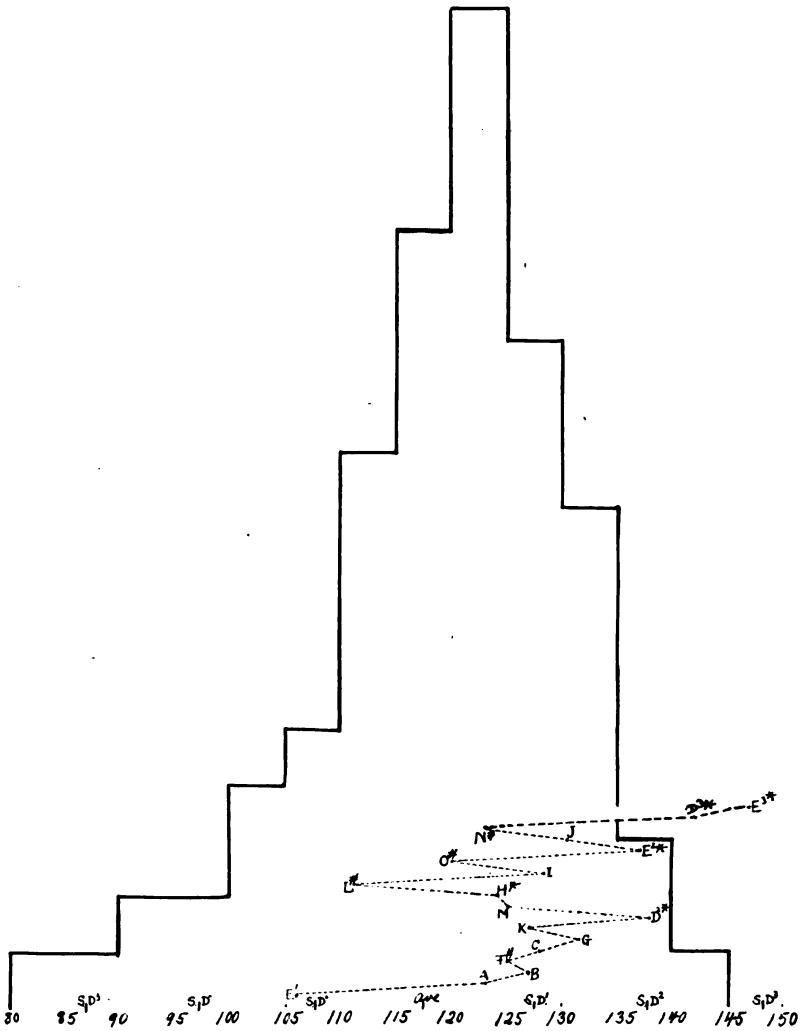


Fig. 21.

A diagram showing the surface of frequency based upon the test scores of the class, and the curve of the successive scores of Subject C. A star (*) indicates the scores achieved at the contraction periods. A double cross (#) indicates the scores at the quiescent periods.

order of tests for Subject H was the reversal of that for Subject C. Their achievements varied similarly and seemed unquestionably to parallel the degree of activity of the stomach. The two indication marks, a star (*) and a double cross (#) in Figs. 21, 22 and 23, prove that in all curves the test scores are high at the hunger contraction periods, and low at the quiescent periods.

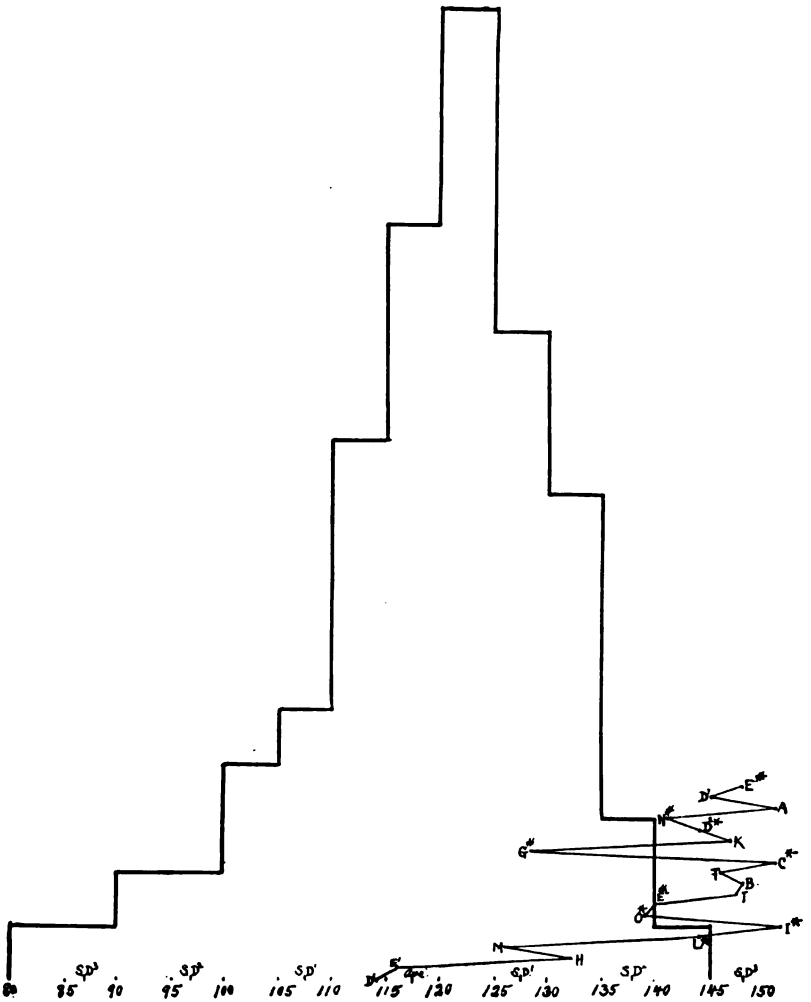


Fig. 22.
A diagram showing the surface of frequency based upon the test scores of the class, and the curve of the successive scores of Subject H. A star (*) indicates the scores gained at the contraction periods. A double cross (#) indicates the scores at the quiescent periods.

The foregoing experiment proved that hunger augments mental efficiency. The effect on fatigue of this continuous mental work appeared in the fact that in the case of Subject C, the hunger contraction periods were much depressed and did not take normal frequency nor vigor, though this was not so marked in the case of H. The subjective feeling of fatigue correlated very closely with the quiescent periods and the low efficiency.

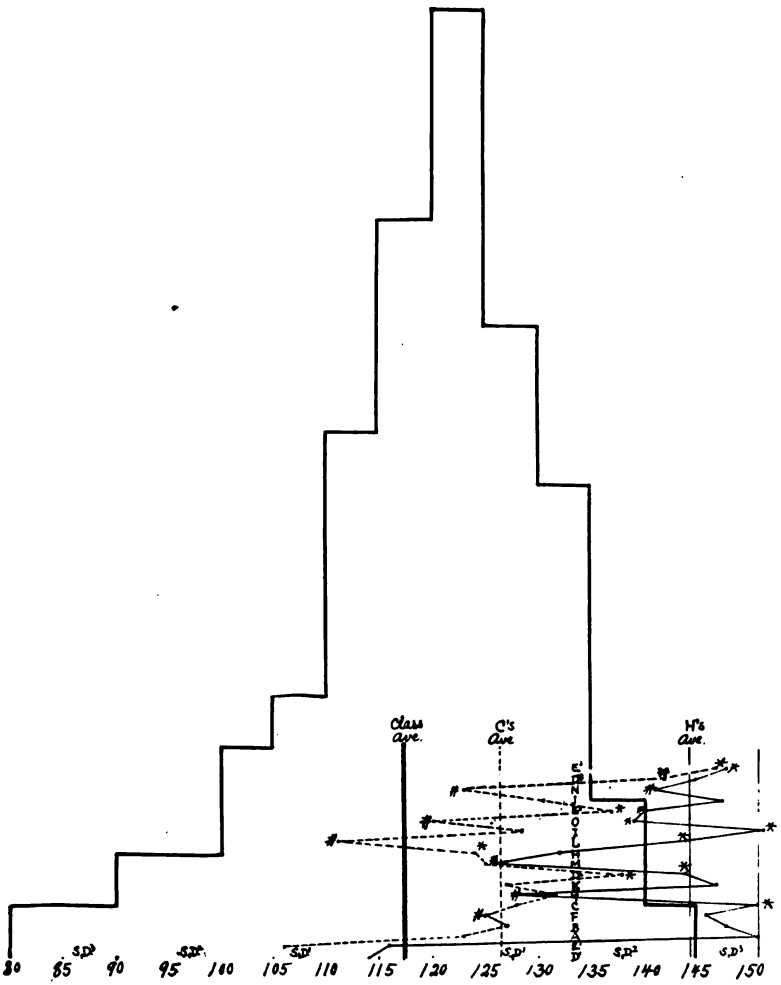


Fig. 23.

A diagram showing the relative position of the scores of two Subjects, C and H, and their averages in the surface of frequency in relation to the class average.

Chapter IX.

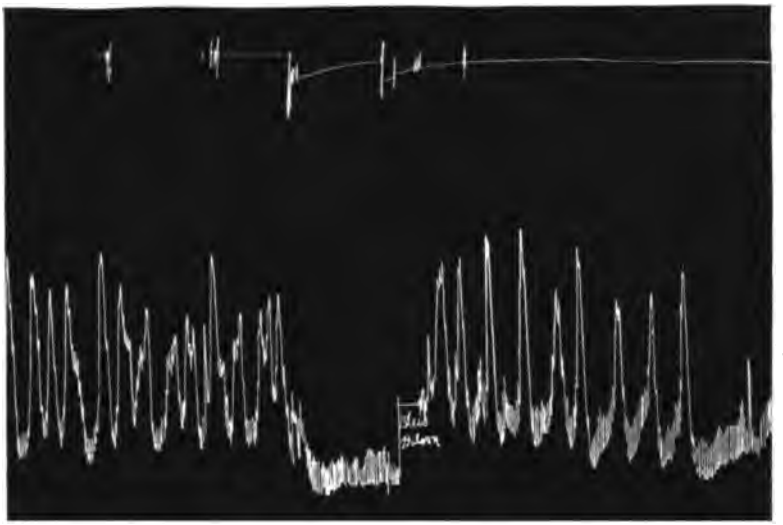
THE EFFECTS OF VARIOUS STIMULI UPON THE HUNGER RHYTHM.

1. The Effect of Mechanical Stimulation.

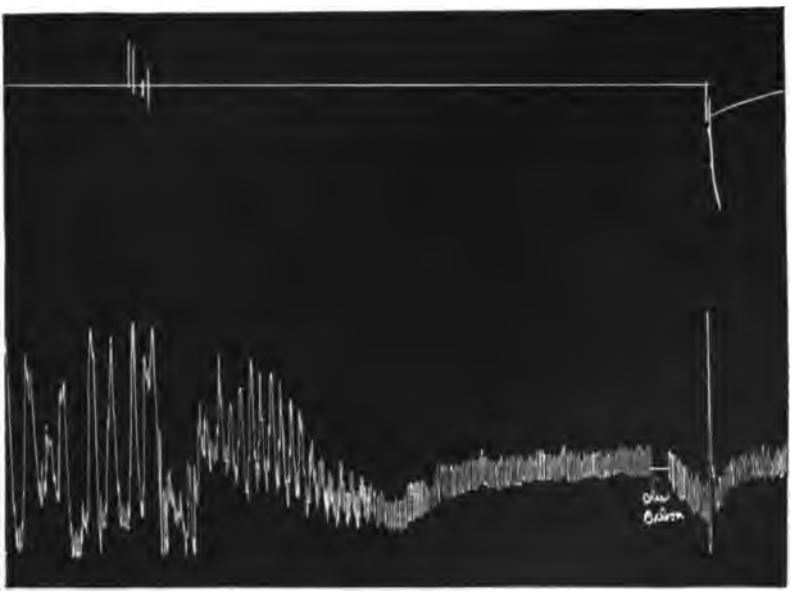
During the night experiment of April 24th the experimenter twice blew the stomach balloon to see if there had been any wreckage, since after strong contraction periods, the respiration pressure rhythm became so low that it extended beyond the limit of the smoked paper. Each time when the tube to the manometer was disconnected for a few minutes and the balloon was distended, there appeared on the record a few stomach contractions, while the stomach was quiescent otherwise. Curiously enough, the bodily movements appeared simultaneously with these contractions. Further study of this phenomenon was taken up on June 9th. At 4:15 A. M., while Subject C was asleep, the stomach balloon was disconnected from the manometer, and was blown up to such an extent as to bring the marker one inch higher in position on the smoked paper. This was done five minutes after a long contraction period, which had lasted twenty minutes. This distention of the balloon was immediately followed by a long contraction period which lasted 45 minutes, the contractions numbering 43. The same stimulus was given at 5:15, ten minutes after this long contraction period. It started one large contraction and a large bodily movement. (See Fig. 24.) At 7:15, ten minutes after a strong contraction period, blowing the balloon produced a one-inch contraction, and a few wrigglings of the body. Ten minutes later, at 7:26, in the midst of a quiescent period, there followed, a few minutes after the stimulus was given, one contraction and a body movement.

On June 10th this experiment was repeated with Subject H. After a strong, one-hour contraction period, a state of quiescence set in, and lasted for 50 minutes. At 6:20, the stimulus of distending the balloon was given. As soon as the apparatus started to mark, it showed a strong contraction period, which lasted for 30 minutes ending with a tetanus.

Fig. 24. A and B. Effects of various stimuli. Mechanical stimulus of blowing balloon.



A.



B.

Subject H was once more tested on the following night, June 11th. There was a contraction period up to 5:20 A. M. which was followed by a tetanus, and a long quiescent period. After the stomach was perfectly quiet for 40 minutes, the stimulus was introduced, at 6:05. This did not bring about any result until five minutes later when there came a gradual start of a contraction period. At 7:20, at the end of a vigorous contraction state, another blowing of the balloon was introduced. Three minutes after there was a contraction and a few bodily movements and a few minutes later the subject awoke.

The above results show that, out of seven experiments, two immediately started contraction periods, lasting 45 and 30 minutes respectively; one started a group of contractions five minutes after the stimulus, and four gave a few contractions and bodily movements. These mechanical stimuli seemed to arouse local reflexes of the stomach, which may be controlled by the plexuses of Auerbach and Meissner. About the same time that the present writer had found these results, Carlson* published his results wherein he concluded that mechanical or electric stimulation of the lung, the gall bladder, the heart, the urinary bladder, and the entire gastro-intestinal tract induces skeletal reflexes both in decerebrated and in purely spinal preparations. The bodily movements that occurred simultaneously with the stomach contraction seemed to prove that the afferent nerve fibres of the viscera make reflex connections with the skeletal motor system, though they may have long latent periods and rapid fatigue, which indicate a complex synaptic system, either in the spinal cord or in the visceral afferent paths.

Carlson assumes that these visceral skeletal reflexes, at least as regards the extremities, are essentially of the defensive or escape type. It suffices, however, to state that the present experiment proved that the mechanical stimulation produced the stomach contractions, as well as bodily movements in man.

2. The Effect of Drugs.

Papaverin, an opium alkaloid ($C_{12}H_{21}NO_4$), has been established, especially on the continent of Europe, as having the effect of paralyzing smooth muscles, such as those of the stomach. Dr. D. Macht, of Johns Hopkins University, has extracted the

* Carlson, A. J.; American Journal of Physiology. Vol. 55, P. 384, 1921.

essence of this drug and produced, by combining it with other solutions, such as alcohol and wine, a little more tasteful drug, called benzyl-benzoate. These drugs are used to stop smooth muscle contractions of the stomach or intestines, in cases of hemorrhage or vomiting. In the present study these two drugs were used to determine (1) how they affected the hunger contractions during sleep, and (2) whether relaxation of the stomach muscles influenced the gross bodily movements.

On April 13th, at 10:30 P. M., Subject B took 4 cc. of benzyl-benzoate by way of the mouth, together with ¼ of a glass of water to clean the mouth. The subject went to bed immediately, and the record was taken, both of the stomach activity and the bodily movement, until he awoke at 8 A. M. next morning. There was a long tonus period, 10:30 P. M. to 1:30 A. M., lasting nearly 3 hours, during which tonus waves were large and distinct. The contraction periods of 1:30 A. M. to 2 A. M. and 3:15 A. M. to 3:30 A. M. were of small amplitude. This experiment showed that (1) the drug gave a primary stimulating effect, so that the tonus waves were of larger amplitude; (2) that the stomach had a longer period of quiescent and tonus states; (3) that the stomach contractions of the early part of the night were of smaller amplitude than usual.

Lest other components of the drug had some stimulating effect, the essence of the original Papaverin was tried on the same subject, on April 20th. From 10:30 P. M. to 2 A. M., the stomach was very quiet. The contraction period from 2 A. M. to 2:30 A. M. was of very small amplitude, and until 5:10 A. M. the stomach was practically atonic. Three contraction periods, 5:30 A. M.—5:45, 6:30—7; and 7—7:30, became gradually larger, and the final one was of the amplitude of 4.5 inches. This result can be interpreted to mean that the drug had the effect of keeping the stomach muscles relaxed until 5:10, and after that, vigorous contractions were resumed gradually. The bodily movements followed quite the same distribution; when the stomach was atonic no body movement could be seen. There was no way of knowing just when the drug had been absorbed, or had begun to have effect. It would be of value to introduce this drug as a subcutaneous injection at the height of each hunger contraction, and determine how long it will have effect in controlling the muscular movements of the stomach.

3. The Effect of Conscious Effort, Thought, Sight, Smell and Taste of Food.

On June 9th, at 6 P. M., when Subject H had contractions, the experimenter asked him to try to stop them. In spite of his effort, large contractions continued. Again he was asked to try to start contractions by his effort. The request was made in the quiescent state, but the subject, of course, did not know it. His conscious effort made no difference on the record. The same request was asked of Subject C on June 10th, at 7:45 A. M., before breakfast. This brought no contraction change, not even respiration change. We can be quite sure, from these results and from other considerations, that the stomach contractions, or quiescence are quite independent of conscious control.

A few minutes before 6 P. M., on June 5th, Subject C was asked to shut his eyes and *think of* delicious dishes—a dish of hot roast beef with mashed potatoes and green peas, salad of fresh lettuce and tomatoes, and hot mince pie—all of which the subject especially liked. This thinking was done while the stomach was quiescent. The record showed no effect. On June 9th Subject H was asked to think of a dish of sweet, juicy and delicious cherries. This thinking was done while the stomach was in contraction, and the record showed no effect. A dish of cherries was then brought in, which the subject was allowed to observe and smell as much as he liked. No change appeared. However, as soon as the food particle was placed on his tongue, salivary secretion gushed out, which, together with some of the juice of the food, was swallowed. The contraction stopped at once. (See Fig. 25.)

In the morning of June 10th, at 7:45, when Subject C awoke from his dream-experiment, the experimenter asked him to think of a dish of sliced oranges for breakfast. This started no contraction, but rather brought down the tonus state into still lower stomach pressure condition. After 5 minutes a dish of juicy sliced oranges was brought in, and the subject looked at them and smelled them. The stomach started no contraction from these stimuli.

It is clear so far (1) that conscious effort to start or to stop stomach contractions is of no effect; (2) that the thought, sight or smell of food has no direct influence to start or to stop the stomach contraction; and (3) that food particles, or juice, or even some amount of saliva swallowed, stops the stomach contraction.

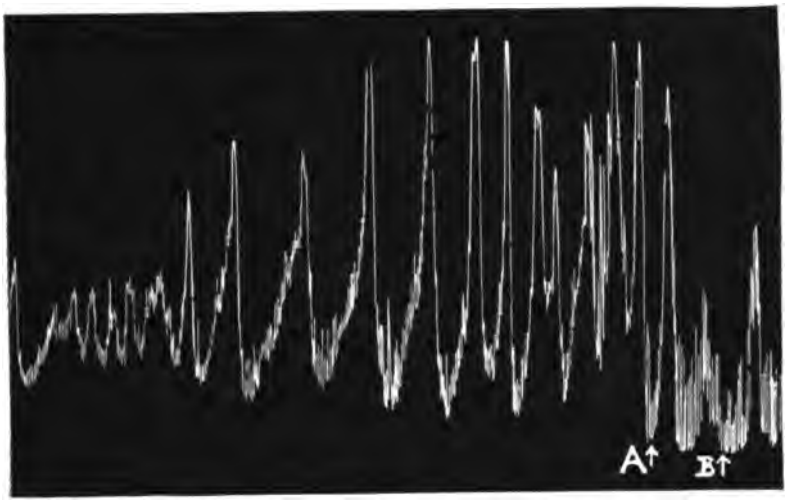


Fig. 25.
(A) Thought of delicious food, hot beefsteak with mashed potatoes and green peas, fresh green salad with tomato slices, and hot mince pie with cheese, did not stop contraction. (B) Actual taste of food stopped contraction and swallowing the juice through the esophagus is marked as if being contraction.

4. The Effect of Nausea from Rotation.

Dr. Knight Dunlap has devised a rotating chair on which a subject sits and can be rotated at the desired speed to the right or left turn by adjusting the switch. On May 22nd, Subject C, who had had no luncheon, sat on this chair. The experimenter waited until the tonus state appeared on the record. Ten minutes' rotation was given at 2:30 P. M. It did not bring nausea, but brought the tonus state of the stomach into perfect quiescence. Thirty minutes later another rotation, 20 minutes long, was given. The subject felt dizzy and nauseated. He looked pale and perspiration was seen on his forehead. The condition of the stomach was absolutely atonic. In spite of the fact that usually this subject had two strong contraction periods in the late afternoon, between 4 and 5, there appeared no hunger period, not even the slightest tonus, up to 5:30. Rotation seemed to have had enough effect to depress the tonus, and the contraction of the stomach. The physiological phenomena of nausea, the general emotional state of disagreeableness, as well as the unfamiliar situation of sitting in an electric chair, may each have contributed to this general depression of the stomach activity.

5. The Effect of Electric Shocks.

On May 21st Subject H came to the laboratory without lunch. He was provided with a wire ring on a finger of the left hand. and a metal plate on the right arm, which were parts of an electric circuit closed by a switch, which was controlled by the experimenter. The experimenter waited until a large tonus appeared before pressing the button of the switch. In the middle of the upgoing tonus rhythm, electric shocks were given. The tonus waves, however, pursued the normal course without being checked by the shock. This was repeated at another tonus period, with the same result. The experimenter expected to try at the contraction periods, but the subject had no contraction period that day. On the following day, Subject C went through the same experiment. The result was similar except that twice, out of ten shocks, tonus waves dropped to the level immediately after the stimulus.

In both cases the subjects felt uncomfortable because of receiving shocks, and they said, "I don't like that!" There was no contraction period in either case. Probably it was this emotional state of general discomfort and fear that influenced and inhibited the occurrence of the contractions. These external electric stimuli seemed to affect the autonomic nervous system, and produce general emotional states, which inhibit the muscular activity of the stomach.

6. The Effect of Prolonged Work.

On April 3rd, at 9 A. M., Subject C began to copy from an Italian book, which he did not understand. He had had no food since the previous evening. He typed with an Underwood machine while the apparatus was recording his hunger states. Up to 1:20 one continuous state of quiescence persisted. After a few minutes of high waves, quiescence followed and lasted for 2 more hours. At 3:30 eleven contractions of 2.2 inches were seen. At 5 P. M. the work was stopped and the subject was allowed to go to sleep. As soon as he fell asleep, hunger contractions of 3½ inches in height appeared. From 5 to 6 P. M. there was a continuous series of strong contractions, which numbered 52 in all. In the normal records, this individual had contraction periods most frequently at 11 o'clock and 1, 2, 4 and 5 o'clock, while in

the work experiment, all the forenoon and afternoon contraction periods were eliminated. The subject himself said that he did not feel hungry at all during the work. Prolonged work, when it involved feelings of fatigue and lack of interest, perhaps through correlation with an emotional condition, tended to inhibit the hunger contractions.

7. The Effect of Reading Exciting Stories.

1. An Interesting Story. On May 9th Subject H swallowed the stomach balloon, and then read "Moon Calf" from 9:30 to 6 P. M. He was alone in an experimental room, and was allowed to use the chair or the couch. He read the story with great absorption, especially the part of the book from the climax on to the end. The record, registered in the next room, showed three contraction periods, which were much shorter and of less amplitude than the normal ones.

2. Stories of Fear and Horror. Edgar A. Poe's stories, "Fall of the House of Usher," "Murders in the Rue Morgue," "The Premature Burial," and "The Pit and the Pendulum" were selected for arousing the emotion of fear and horror in the subject. From 1:45 to 6 P. M. on June 5th Subject C read these stories. In spite of the fact that he had had nothing to eat since morning, the stomach showed no contraction until 4:30, when very weak contraction waves started. On the other hand, deep breathing, which habitually occurred once every two or three minutes, in this individual, was so much deeper that, on the record, it measured twice the amplitude of the usual ones.

3. Sad Stories. On June 9th, at 9:50 A. M., the experiment was started with Subject C, who had had nothing to eat since the previous evening at supper. The selected stories were "Vanka," by Turgenief, "Hide and Seek," by Sologub, and "The Death of Nell" from Dickens' "Old Curiosity Shop." The record showed that there was general depression of the stomach activities. No contraction period appeared in the three-hour experiment.

Thus far the results showed that certain types of reading brought about diffused, non-discriminable, and general emotional states, which tended to inhibit the full swing of the automatic movements of the stomach.

CHAPTER X.

CONCLUSIONS.

The present experimental study brought to light a certain number of facts that show the extent to which hunger plays its rôle as a dynamic force in the activities of life.

The primary cause of the physiological and psychological phenomenon of hunger is ascribed to a certain muscular contraction of the stomach. The hunger contractions of the stomach occur from 3 to 4 hours after a meal and recur at intervals as long as the stomach is empty. The start of the hunger contractions is controlled by the intrinsic nerves and is independent of the extrinsic nerves which have only regulatory effects. Thus the stomach is a self-starting organ and yet has intimate connection with other vital organs and glands, through the autonomic nervous system which maintains the general tone of the organic background of emotional and perhaps intellectual life.

The effect of the gastric hunger contractions upon other physiological conditions was investigated as reported in Chapter II. Changes in respiration, vasomotor flow, and salivary secretion were studied in relation to the hunger contraction rhythm. Especially close correlation was found in the case of the salivary flow, which is one of the most immediate accessory phenomena of the hunger contractions.

That the sensation of hunger occurs synchronously with the hunger contraction of the stomach was assured in the present experiment with four subjects, men and women, normal and clinical. This fact may be interpreted as showing that the stimulus from the gastric organ affects the cortical integration centers in some way so as to set up impulses toward the co-ordinated action of securing food.

The stomach contractions not only arouse the sensation of hunger but prepare the whole body for activity, even during unconscious states. The experiments on bodily movements of men, infants, and rats, showed that bodily movements occur simultaneously with the hunger contractions, while in the quiescent periods very few bodily movements occur. This was especially marked when the subject was asleep. In the cases of the infants and rats the periodical recurrence of bodily movements was strikingly regular.

In sleep we do not know to what extent the visceral stimuli of the hunger contractions affect the central nervous system. The states of dreaming and non-dreaming, however, seem to indicate something of the mental status of a sleeping man. The investigation of the correlation of the hunger contraction periods with the dreaming periods revealed it to be positive. Though we are sure that there are also other factors that control the occurrence and nature of dreams, the present results warrant the statement that men dream more at the hunger contraction periods than during quiescence.

The augmentation of motor activity was tested by means of a hand-dynamomoter which was found to be less subject to the effects of practice than other familiar motor tests. The results showed that at the hunger contraction periods the power of grip is greater than at the quiescent or after dinner periods. The average of all contraction period scores compared with those quiescent and after-dinner-period scores is 88.98 kg. ($P.E._{A.} = .55$) vs. 86.64 ($P.E._{A.} = .49$) and 81.84 ($P.E._{A.} = .40$) for Subject C and 90.91 kg. ($P.E._{A.} = .22$) vs. 87.11 ($P.E._{A.} = .28$) and 86.42 ($P.E._{A.} = .40$) for Subject H.

A further experiment to measure the effect of the hunger contractions in its relation with mental activity, Chapter VIII, produced the following facts: using fifteen forms of the Thorndike Intelligence Examination, Part I, a high correlation between the hunger contraction periods and test scores was found. These tests are relatively free from the effect of practice, but vary according to the presence or absence of the hunger contractions. It seems safe, therefore, to say that hunger augments efficiency in mental work.

Thus far the experiments were directed toward measuring the effect of hunger upon physiological conditions, the sensation of hunger, bodily movements during waking and sleeping states, motor activity and mental activity. The next group of experiments had the purpose of investigating the effects of mechanical, chemical, conscious, and emotional stimuli upon the hunger states.

The mechanical stimulation of distending the balloon in the stomach produced a group of contractions accompanied by bodily movements of the sleeping man. The chemical action of drugs, benzyl-benzoate and papaverin, which have relaxing and tonus-lowering influence on the smooth muscle structures, depressed the periodicity and vigor of the hunger contractions. Less frequent and less vigorous appearance of the hunger contractions resulted in fewer and weaker movements of the body of the sleeping man.

The conscious effort to start or to stop the contractions of the stomach had no influence in controlling the hunger rhythm. The thought, sight, smell, and taste of appetizing food failed to have any observable affect upon the gastric hunger states. These facts prove that the hunger rhythm is not a matter that can be controlled directly by the conscious center of "will" or motive.

The nervous system that regulates the tonus and the rhythm of hunger contraction is very sensitive to emotional stimuli. The effect of nausea from rotation, electric shocks, and fatigue from prolonged work, showed that as the consequence, the hunger rhythm is distorted and the contractions are inhibited. Reading exciting stories brought about emotional situations which, irrespective of whether they were of interest, fear, horror, or sorrow, had inhibitory effect upon the hunger rhythm. Emotional stimuli, therefore, affect the autonomic (especially the sympathetic) nervous system which has an inhibitory influence upon the visceral function of the stomach.

From the preceding conclusions of the present experiment it is evident that hunger is not merely a local function of a local organ. It involves the whole organism, and it stimulates the motor and mental apparatus so as to bring the organism into a state of readiness to secure food. Biological consideration of the hunger behavior of lower animals, permits us to say that the state of tension of the whole organism prepares it for the strife and struggle against the enemies which lie between it and its food.

Hunger is ontogenetically and phylogenetically one of the oldest phenomena, appearing in the first day of the infant life and in the first days of the unicellular ancestor. The hunger mechanism is one of the most primitive organs, in the sense that its activities are carried on partly by automatic activities of the organ without extrinsic nervous control, and partly by the sympathetic nervous system, which is said to be "a direct survival of that diffused type of nervous system which alone is found in the lowest animals which possess nerves at all, such as some jelly fishes and worms. It serves to supplement the non-nervous protoplasmic activities of the different tissues which co-operate in the performance of the work of the several organs."* Yet this most primitive mass of tissues and nerves forms the vital organ for the survival of its owner, and gives moreover a characteristic organic background to the entire conscious life. What Herrick* expressed in connection with emotions in general is found here to be applicable

* Herrick, C. Judson; An Introduction to Neurology. 1918, P. 252, 290.

to hunger also. "In normal man the mechanisms may function with a minimum of cortical control, giving the general tone of well-being or malaise, or they may be tied up with the most complex cortical processes, thus entering into the fabric of the higher sentiments and affections and becoming important factors in shaping human conduct."

Thus hunger determines not only the inner conditions of man, but works as a driving force for the projection apparatus, arms and legs, etc., to act upon the outer world. When viewed sociologically we are amazed at the fact that the most complicated events may be traced back to the simple and normal craving for food. In the economic world, food stuff is the barometer for the rise and fall of the cost of living.

The fact that hunger can not be voluntarily controlled, does not mean, however, that there is no way of guiding and educating this fundamental force of human life. The laws of adaptation established in the studies of conditioned reflexes, modified reactions to stimuli, substitution, and sublimation find their application here also. Especially in early childhood, educators find opportunities to direct the habit formation of infants according to the law of hunger and its periodicity, the law of sleep and activity periods in relation to hunger, instead of according to tradition and haphazard ways. For the problems of work periods in school and other occupations, if the leaders keep in mind the same fundamental laws, they can utilize and co-operate with these natural tendencies for activity and non-activity periods. Without the realization of this biological readiness of the organism in hunger which has been hitherto called the instinct of food-seeking, short-sighted guidance and suppression may give rise to riots, revolutions, and international wars. The degree to which we are able to control the dynamic force of life depends upon the extent of our knowledge of the force.

INDIVIDUAL DIFFERENCES AS AFFECTED BY PRACTICE

BY
GEORGINA STICKLAND GATES, Ph. D.

ARCHIVES OF PSYCHOLOGY
EDITED BY R. S. WOODWORTH
No. 58

NEW YORK
AUGUST, 1922

AGENTS: G. E. STECHERT & CO. : London (2 Star Yard, Carey St., W. C.); Paris (16, rue de Condé)

TABLE OF CONTENTS

CHAPTER		PAGE
1.	Problems and Methods of Investigation	5
2.	Changes in Individual Scores and Group Relations with Practice	10
3.	Inter-correlations of the Tests	19
4.	The Relation of Improvement to Initial Ability	37
5.	The Relation of Improvement to Final Ability	63
6.	The Relation of Improvement in one Test to Improvement in Another and of Initial and Final Gain	68
7.	Summary	72

ACKNOWLEDGMENT

I am indebted to Professor Thorndike for a number of suggestions as to statistical procedure; to Professor Woodworth for advice concerning changes to be made in the presentation of the study; and to Professor H. L. Hollingworth for the original problem, for the opportunity to conduct the experiments, for the use of valuable records, and for constant interest and encouragement.

This paper was completed, aside from some revision of the text, before October 1, 1919.

Individual Differences as Affected by Practice

CHAPTER ONE

PROBLEMS AND METHODS OF INVESTIGATION

In the literature of psychology, a large number of studies have been included under the term "practice experiments." Investigations in this field range from studies of learning Russian or telegraphy to inquiries concerning improvement in ability to tap. Experiments have been performed upon children; upon adults under normal conditions, as when working in a psychological laboratory; upon adults in various unusual conditions as when dosed with caffein or alcohol, or working in a humid atmosphere; observations have been made of the effect of the repetition of a situation upon monkeys, dogs, cats, chickens, turtles, paramocium. Such problems as the determination of the factors in learning, of the typical course of improvement, of the adequate representation of learning, the necessity for and causes of plateaus, and the question of the probable limit of improvement and of the frequency of approach to such a limit, have been attacked many times both by theoretical analyses, and by the method of "patience, starving out, and harrassing to death."

The majority of such investigations, whether mainly theoretical or mainly practical in their implications, are concerned with changes in individual scores. Attention is directed to the subject's initial and final status and to variations observed in his measured efficiency. On the basis of such objective records, an attempt is made to determine the factors responsible for improvement, retrogression or for apparent lack of change. By varying conditions, the influence of external and internal events upon the effectiveness of the repetition, as determined by the score, are studied.

Such investigations deal mainly with phenomena in which individuals are assumed to resemble each other. As we assume that all children of a given physical age will vary about a norm, just as do a certain random sampling—say 500—children of that age, so it may be suggested that most people's

learning of telegraphy will resemble, at least in important factors, the amount, rate, and kind of improvement found in the five persons studied. Discoveries concerning the effect of practice have been, in the main, results obtained from careful experimentation on a few persons. In typical experiments, for example, Romer, Weygant, Vogt, Kapemen, and Swift, had each one subject; Rejall, Hill and Book had two; Bair had three; Coover and Angell four; Bryan, Harter, Swift, Heiman, and James had five; Thorndike, Starch and Fracker had eight; Whitley had nine; Wells, Hahn and Dearborn had ten; Kline, Leuba and Hyde had seventeen; Munn had twenty-three. We have a few studies like those of Kirby who had seven hundred; Brown and Peterson with thirty-four; Thorndike with forty-three. For the most part the number of subjects has been small, and in those cases where a large number of individuals were studied, the functions tested were usually but one or two.

The interest in the results drawn from the individual scores of a few persons studied intensively was in the improvement phenomena itself, rather than in any further problem. The authors inquire what is learning; under what conditions does it come about; what factors are important for improvement; how may we facilitate it, rather than what effect does the knowledge gained from practice experiments have upon some other psychological concept.

The present study differs from the classical investigations in that it is not concerned primarily with individual scores but with relations between individual records, in that individual differences rather than individual similarities are stressed, and in that the practice phenomena are studied only with a view to the light which they may throw upon a further question—that of the nature of native or acquired differences between the subjects of the investigation.

For this reason, we inquire whether the relative position of individuals in the group remains constant or changes with practice; whether there is any relation between proficiency in the various functions before and after opportunity for improvement; we attempt to discover the factors which determine initial and final status; and the relation between gains made in different tests or in different parts of the practice period.

The main questions in the order in which they are discussed are:—

1. What changes in individual scores occur as a result of the repetition of a performance?
2. Does the variability of scores decrease with continued practice?
3. Do group relations remain unchanged after repetition? Does the best person remain at the top and the worst person at the bottom? Is the first trial diagnostic of the succeeding trials and how well does the last trial correlate with those preceding?
4. What is the correlation between the functions? Does this change with practice?
5. What is the correlation between initial ability and improvement?
6. How are final capacity and amount of gain correlated?
7. Is gain in one test indicative of gain in another?
8. Is gain at the beginning of the practice period diagnostic of gain toward the end?

Method

The following table gives the tests used and the method of scoring.

TESTS AND METHOD OF SCORING

Color-Naming: Two Woodworth-Wells blanks pasted together. Record, time required to name correctly a set of two hundred colors. Test repeated twenty-five times.

Tapping: Record, number of taps executed in two minutes with the hand stylus, right hand. Test repeated twenty-five times.

Adding: Kraepelin blank. Record, time required to add one hundred examples of two numbers each. Test repeated thirty times.

Multiplying: Record, time required to solve mentally a set of five multiplication examples. The examples required the multiplying of one two-place by another two-place number. No number below four was used. Different examples were used each time up to the twentieth trial. The twenty-first and twenty-second sets were repetitions of the first and second sets of examples.

Word-building: Record, number of shorter words built from a given word (as weather, psychiatry, etc.) in a specified time. Words varied in difficulty and time allowed was changed as the experiment proceeded. Twenty different words used. The

twenty-first and twenty-second words were the same as the first and second words.

The tests, it will be noted, vary from the very simple tapping performance in which we may assume that practice means the exercise of relatively few bonds, to the more complex tests such as multiplying and word-building in which the subject's reaction, performed each time in response to a different problem, is the result of a comparatively elaborate cerebral process. The tests differ also in their content, and in the method of scoring. One record is concerned with a motor performance; two have to do with words, and two with numbers. Three functions are scored by the time and three by the amount method. An equal number of trials of color-naming, adding and multiplying means, then, an equal number of trials successfully performed regardless of time spent. Practice trials of tapping and word-building, on the other hand, refer to definite temporal periods.

The subjects were twenty-three women students taking their second or third terms work in psychology. The group, though not more homogeneous than groups used for similar experiments, were much more alike in age, experience, interest, and intelligence, than twenty-three people chosen at random.

All the tests, except word-building were given to the subjects individually, by one experimenter. Each subject came to the experimenter for two separate half-hours each week. The time of day, exact interval between test periods varied, of course, for different subjects, but for any one individual, the time and regularity of performance were kept as constant as possible.

For the final results, twenty-nine trials of the adding test and twenty-two trials of each of the other tests were used. All computations are based on a score which is either the first trial, the very best trial, the median of three, or eleven, or (in adding) four trials, or the best of three or four trials.

For the correlations, the formula $\rho = 1 - \frac{6\,\Sigma d^2}{n(n^2-1)}$ was used. ρ has not been changed into r.

A table of probable errors of coefficients computed for twenty-five individuals, and one showing the divisions of the prac-

tice period which will be used in chapters two and three, follow:

*Table of Probable Errors**

Using Formula: $P.E. = \frac{.67\ (1 - r^2)}{n}$

r	Group n = 23
.1	.14
.2	.14
.3	.13
.4	.12
.5	.10
.6	.09
.7	.07
.8	.05
.9	.03

The Divisions of the Practice Series

1st Division. The initial trial.

2nd Division. The second, third, and fourth trial of color-naming, tapping, multiplying, word-building. The second, third, fourth and fifth trial of adding.

3rd Division. Fifth, sixth, seventh of four tests. Sixth, seventh, eighth, ninth of adding.

4th Division. Eighth, ninth, tenth of four tests. Tenth, eleventh, twelfth, thirteenth of adding.

5th Division. Eleventh, twelfth, thirteenth of four tests. Fourteenth, fifteenth, sixteenth, seventeenth of adding.

6th Division. Fourteenth, fifteenth, sixteenth of four tests. Eighteenth, nineteenth, twentieth, twenty-first of adding.

7th Division. Seventeenth, eighteenth, nineteenth of four tests. Twenty-second, twenty-third, twenty-fourth, twenty-fifth of adding.

8th Division. Twentieth, twenty-first, twenty-second of four tests. Twenty-sixth, twenty-seventh, twenty-eighth, twenty-ninth of adding.

*This formula, it should be noted, is strictly applicable only to the Pearson coefficients.

CHAPTER TWO

CHANGES IN INDIVIDUAL SCORES AND GROUP RELATIONS WITH PRACTICE

Before we can ask what changes relations between individuals undergo as a result of practice we must know exactly what effect repetition in the present case has upon individual scores. At the risk of appearing to prove the obvious, we must assure ourselves of the existence, amount and relative rate of improvement in these tests.

We may ask what transformation takes place in the average of the group as a whole. Does the average time required decrease or increase? Is the average amount of work performed larger or smaller? This information, though valuable, is not sufficient. It does not tell us enough about individual improvement or retrogression. A small amount of gain in the average of the group might mean a large improvement on the part of some persons and a stationary position or retrogression by others. In order to tell the story adequately, it might seem necessary to present all the records in full, but as such tables would be cumbersome and, as all the correlations in Chapter Three and Four are based upon median or best scores, be unnecessary for a correct interpretation of the facts. Here is included a table (Table 1) showing the group average and the individual scores for the first trial and the group average and individual median for the other seven divisions (see previous Chapter) of four of the tests.

Because of the varying difficulty of the words used and the change in time allowed for word-building, it is difficult to obtain from this test information as to amount of improvement, change in rate, or approximation to a practice level. For this reason it is omitted here.

If we consider first group averages—found in the last line of Table 1—we see that there is gradual improvement in all tests. The average of the adding improves up to the sixth division, color-naming to the seventh division, multiplying and tapping improve throughout the practice period. The improvement in tapping and color-naming is relatively small, in multiplying and adding there is a relatively large gross

amount of improvement. In tapping the improvement is approximately one-twelfth of the final score, in color-naming it is one-fifth, in adding one-half and in multiplying about twice as great as the final record.

The gross gain as measured is greatest at the beginning of the practice series.* From the first to the third periods there is a change in averages of 12.6 seconds in color-naming, 45.3 taps in tapping, 30.2 seconds in adding, 232 seconds in multiplying. Between the sixth and eighth periods there is a change of only 8 seconds in color-naming, of 7. 4 taps in tapping, a gain of 8 seconds in adding, a drop of 59 seconds in multiplying.

Adjacent differences are given in the following table:

TABLE NO. 2

Adjacent Differences in Average Scores.

	Δ1-2	Δ2-3	Δ3-4	Δ4-5	Δ5-6	Δ6-7	Δ7-8
Color-Naming	9.7	2.9	.5	3.1	.8	1.6	8
Tapping	18.7	26.6	–.7	11.3	6.3	5.6	1.8
Adding	21.1	9.1	2.3	.5	2.8	.5	.3
Mult.	214	18	15	29.2	23.7	12.5	46.1

The largest difference is (with the exception of tapping) between the initial trial and the median of the next three. All but multiplying show a very small or entire lack of improvement at the end.†

Inspection of individual scores reveals the same phenomenon —a relatively large amount of improvement at the beginning which becomes less as practice continues. At the end of the period there may be little or no change in score. The particular question of individual progression will be discussed in Chapters Four and Five. The individual scores, presented here, we need only observe bear out, in general, the results noted from a comparison of average cases.

The changes observed in individual records may or may not produce a change in the relative position of the subjects. A second question is, consequently,—Do persons improve uniformly with reference to one another? Does the best person tend to stay at the top, the worst person at the foot, the me-

*Our method of scoring may, of course, be inaccurate. It may be that gains made at the end of the series should be given a higher score. The improvement curve may really be a straight line. This is not questioned here. Reference is made only to gross measured scores.

†The larger improvement noted in multiplying may have been due, not only to the fact that the function was a more complex one with more opportunity for improvement, but to the repetition at the end of the practice series of the first two multiplication examples. All subjects denied the slightest memory of these examples. Yet the large amount of gain noted there may be due to some subliminal familiarity.

diocre individual in the middle range? Is a record made in an initial test at all indicative of what ability would be at the end of twenty-two trials? Correlating the first trial with the medians of the other seven divisions, we obtain the following table. In this it is possible to include word-building, for changes in difficulty of words or in time spent, as they are constant at each trial for all persons, do not affect conclusions based upon relative position.

TABLE NO. 8

Correlation—First Trial with Median of Other Seven Divisions in Same Test.

	2	3	4	5	6	7	8
C. N.	78	74	85	75	77	79	72
Tap.	62	39	53	45	53	29	37
Add.	45	36	45	48	44	49	33
Mul.	42	45	53	57	50	33	67
W. B.	82	85	83	75	70	81	72
Average	62	56	64	60	59	54	56
Ave., excluding multiplying	67	59	67	61	61	60	54

There is a shift in relative positions as a result of practice. But increasing the distance between the compared trial does not decrease intercorrelations as much as one would expect. The median of the second, third and fourth trials resembles more closely (with the exception of multiplying)* the initial record than does the median of the twentieth, twenty-first and twenty-second repetitions. But the difference is slight, with the exception of tapping and possibly word-building—well within the probable error of the average coefficient. Averages in fact are practically identical. There is a suggestion that correlations do decrease but the difference in figures is too small and the appearance of high coefficients too variable for any proof.

A correlation of the median records in the second division with the median records of the other six divisions gives similar results. There, coefficients are all, as we would expect from their increased reliability, a little higher than those in Table No. 3. The correlations given by tapping and multiplying decrease. The others are practically unchanged.

When the median of the last division is correlated with the medians of the other divisions the coefficients found are, in general, higher than those obtained in Tables Three and Four. This may be explained as due to the increased reliability of the last trial.

*The final high correlation of multiplying may be due in part to the repetition at the end of the series, of the first two sets of examples. This does not seem to affect word-building.

TABLE NO. 4

Median of Second Division Correlated with Median of Other Six Divisions.

	3rd	4th	5th	6th	7th	8th
Color-naming	90	89	88	92	87	90
Tapping	72	79	78	71	54	63
Adding	85	89	89	86	74	82
Multiplying	64	50	55	35	19	43
Word-building	79	85	75	68	73	77
Average	78	78	77	70	61	71

TABLE NO. 5

Correlation of Last Division with Other Divisions.

	1st	2nd	3rd	4th	5th	6th	7th
C. N.	72	90	89	85	91	94	94
Tap.	37	63	92	95	92	94	86
Add.	33	82	70	80	84	84	86
Mul.	67	43	68	82	80	80	74
W. B.	72	77	80	79	75	80	86
Ave.	56	71	80	84	84	86	85

The coefficients of tapping, adding and word-building increase. Multiplying is irregular and color-naming, after the first division is uniformally high. The average does not increase after the fourth division. A correlation of .85 or more is reached in the color-naming test in the second division, in tapping in the third, in adding and word-building in the eighth, in multiplying in the second. Evidenty four trials of a color-naming test will give a fairly reliable index of an individual's final capacity as displayed after 22 trials of the test. Seven are needed for tapping, twenty-two for adding and word-building. The correlation of multiplying varies so much that it is unreliable.

Similar correlations have been found by Whitley, Wells, Chapman, Brown and Hollingworth.* The decrease in Hollingworth's result is marked. The small size of many of the coefficients and the changes observed in the course of the experiment are attributable, we believe, to the great length of the practice period. Individuals, in this experiment, did actually reach their limit of improvement. The table is repeated on page 14.

In the present experiment it seems as though individuals maintained their positions fairly well, though there is a definite amount of change. In some cases the difference increases as the distance between repetitions is increased. This appears most plainly in Table No. 5 where the most reliable

*Hollingworth "Individual Differences Before, During and After Practice."

TABLE NO. 6

Correlation of Order of Position of Thirteen Individuals Before, During and After Practice.

The correlation is in each case with the final order, after 175 practice trials (in two cases 130 trials).

Test	Preliminary Trial	5th Trial	25th Trial	50th Trial	80th Trial	130th Trial	175th Trial
Adding	154	193	874	869	973	962	1,000
Opposites	088	616	490	835	945	984	1,000
Colors	682	891	858	913	968	968	1,000
Discrimination	676	621	604	500	500	785	1,000
Cancellation	665	676	885	686	934	1,000	....
Co-ordination	528	793	770	902	946	1,000	....
Tapping	231	484	627	682	693	885	1,000
	41	61	73	79	85	92	1,000

score (final record) is used as the basis for comparison. We might attribute lack of higher correlation between initial scores to unreliability of the determination. Correction for attenuation would probably raise initial coefficients more than final so the difference, presumably, would appear more plainly. On the other hand, as the figures stand, in many cases correlations between first and last records vary only as they might by chance from correlations between first and second division scores.

Beside observing changes in individual scores and in position as compared with initial rank, it is necessary to note the relative variability at the beginning and end of the practice period. How reliable are the records which we have selected as measures of individual ability? Is there any change in reliability of scores with increased practice? We have determined the reliability of the first trial, of the median of the second or third division, of the median of the fourth or sixth division, and of the median of the seventh or eighth division. The reliability of the initial trial is obtained by correlating it with the median of the next three trials. The reliability of the medians of the second or third division is obtained from the adjacent correlation between the two. The correlation between the fifth and sixth, and seventh and eighth medians likewise gives their reliability.

TABLE NO. 7

Self-Correlation of Median Records.

Test	Cor. 1 & 2	2 & 3	4 & 5	7 & 8
Color-naming	78	90	95	94
Tapping	62	72	91	86
Adding	45	85	95	86
Multiplying	42	64	68	74
Word-building	82	79	75	86
Average	62	78	85	85

The median of the last three trials has on the average a higher correlation of adjacent medians than has the median of the second and third trials, which is similarly more reliable than the initial trial alone. In the simpler functions, trials near but not at the end of the practice period (divisions 5 & 6) are the most reliable. Considering the first trial only the order of reliability of the tests seems to be word-building, color-naming, tapping, adding and mutiplying. An observation of the last trial shows that color-naming is the least variable, word-building, adding and tapping come next and multiplying, which shows throughout a lower self-correlation, last.

Another measure which will be employed for certain computations, is one which we have called "best" records. Instead of taking the median of three measures in a practice division as a measure of individual ability as observed at that time, we have taken the highest of these three records. Such a score might serve to eliminate various sources of error (as distraction, temporary lack of interest, fatigue) which would operate more frequently to lower than to raise individual records. On the other hand a "best" score, since it is a determination based upon only one record, might be raised above its actual amount by some chance factor, an error made in recording or some "accident in the nervous system." If such a measure is to be employed, the reliability of the scores must be determined and a comparison of the variability of median and of best records is valuable. Self-correlations of best records at the beginning and at the end of practice obtained in the same way as were self-correlations of median records give us the following table.

TABLE NO. 8

	2 & 3	7 & 8
C. N.	90	96
Tap.	72	94
Add.	81	89
Mul.	59	64
W. B.	73	71
Average	75	83

With the exception of word-building the correlations increase. The final figures for color-naming, tapping and adding are slightly higher than corresponding figures for median records, multiplying and word-building have each a lower correlation in best records as compared with median scores. On the whole, the best record is fairly reliable. There is scarcely more or less variability than there is in the median

record. The more complex functions show a slightly greater regularity when measured by median records, the simpler functions seem to vary at the last trial less when scores are computed in terms of best records.

The facts here suggested are in agreement, in general, with those found by other students of practice. We have found an improvement after repetition, larger in some functons than others, greater in all cases at the beginning than at the end of the period of repetition. This decrease in improvement of individual scores is paralleled by a decrease in amount of change made in the relative position of the individuals. Where scores are highest (about the sixth or seventh periods) reliability is greatest. Where the largest amount of improvement through practice is taking place—as at the beginning of the series—there is the largest amount of variability in relative position. First scores in some cases resemble less and less closely other records, as the number of intervening repetitions is increased. Practice in these tests, then, makes for an increase in scores. With this increase there goes a more or less slight shift in relative position, which is greatest where individual scores and group averages show the largest amount of gain.

TABLE NO. 1

Median Record for Each Individual in Each Division of the Practice Series.

Color-Naming.

Divisions Individual.	1	2	3	4	5	6	7	8
A	117	114	113	120	110	106	107	101
B	098	88	90	89	91	91	85	84
C	203	137	146	149	143	132	138	140
D	122	123	110	120	114	117	107	111
E	113	115	109	97	107	111	103	105
F	100	101	101	99	101	96	99	108
G	105	95	103	100	97	94	95	96
H	110	108	95	96	94	97	87	98
I	075	73	73	72	68	70	66	69
J	123	93	86	98	88	78	80	80
K	110	107	100	98	95	94	88	91
L	096	90	98	90	92	89	83	89
M	095	96	97	92	91	85	82	83
N	124	122	118	108	108	105	104	107
O	115	93	97	96	93	93	96	92
P	097	94	88	95	87	86	87	91
Q	103	89	84	91	88	89	82	82
R	110	97	100	95	92	92	89	93
S	117	117	106	115	98	99	97	102
T	133	122	110	112	99	108	110	108
U	132	114	110	112	101	106	107	106
V	111	105	98	103	102	96	59	60
W	084	76	70	67				
Av.	112.7	103	100.1	100.6	96.5	95.7	94.1	94.9

TABLE NO. 1
Tapping.

Divisions Individual	1	2	3	4	5	6	7	8
A	686	748	706	747	743	748	764	792
B	720	805	766	797	803	794	825	801
C	708	716	635	680	635	668	689	708
D	699	828	806	748	795	718	785	832
E	745	654	654	639	610	610	604	645
F	845	886	826	838	167	862	880	896
G	863	895	943	1001	1104	1087	1040	1051
H	850	882	950	965	918	926	870	904
I	824	825	983	905	879	895	815	930
J	910	888	912	855	927	904	831	882
K	769	740	762	840	756	789	756	769
L	843	748	788	792	795	811	824	803
M	755	787	794	807	801	912	848	811
N	665	645	790	639	686	685	709	668
O	770	791	737	810	925	934	931	919
P	735	749	780	741	767	735	726	727
Q	714	670	684	710	758	807	870	834
R	615	710	824	846	845	860	898	931
S	728	822	820	855	840	860	850	788
T	725	695	764	737	716	675	722	700
U	713	730	787	790	761	734	870	834
V	686	686	675	675	690	695	704	715
W	692	795	919	872	1028	976	912	924
Average	750.6	769.3	795.9	795.2	806.5	812.8	818.4	820.2

TABLE NO. 1
Adding.

Divisions Individual	1	2	3	4	5	6	7	8
A	103	97.5	89.0	82.5	83.5	83.5	84.0	80.0
B	124	87.0	76.5	70.5	72.5	64.4	66.5	69.5
C	118	94.5	92.0	87.5	89.0	84.0	81.5	82.0
D	094	83.5	78.5	75.5	73.0	73.0	73.5	73.5
E	090	85.5	75.5	74.0	70.5	69.0	69.0	75.5
F	097	98.5	89.5	92.0	90.5	84.0	70.5	80.0
G	129	88.5	76.5	76.5	73.5	70.5	71.0	72.0
H	094	88.0	78.0	77.0	78.0	74.0	72.0	79.0
I	101	89.5	75.0	78.0	77.0	71.5	76.5	73.0
J	077	82.5	71.5	65.0	65.0	64.0	63.0	64.5
K	132	98.5	85.0	83.0	80.5	77.5	76.0	79.5
L	084	70.5	63.5	61.5	56.5	56.0	55.5	55.5
M	096	89.5	79.0	74.5	75.0	69.0	72.5	74.0
N	102	76.5	68.5	67.1	66.5	64.0	67.0	64.5
O	094	74.5	69.5	67.0	61.5	63.0	65.5	60.5
P	122	79.1	78.1	78.1	75.6	76.3	82.3	81.6
Q	101	66.1	63.5	64.0	67.1	62.4	62.0	61.7
R	103	75.1	68.6	62.9	62.4	61.2	65.3	63.4
S	115	85.0	74.6	70.9	73.9	70.3	71.2	65.9
T	113	78.0	64.5	62.0	64.3	63.1	62.4	59.5
U	142	94.0	77.5	80.1	80.5	75.3	76.0	76.3
V	112	88.0	77.0	73.3	71.8	71.3	73.0	71.5
W	82	69.0	59.1	54.9	56.1	53.8	56.8	57.1
Average	105.4	84.3	75.2	72.9	72.4	69.6	70.1	70.4

TABLE NO. 1
Multiplying.

Divisions Individual	1	2	3	4	5	6	7	8
A	303	538	370	282	223	200	215	152
B	665	485	320	310	255	270	231	199
C	1104	360	393	447	376	374	358	321
D	730	268	217	239	203	290	172	177
E	613	474	203	182	266	199	169	151
F	877	311	405	735	439	360	524	424
G	480	316	376	238	272	263	263	165
H	374	371	406	259	230	180	231	166
I	778	707	504	554	539	522	429	373
J	541	405	334	198	244	214	227	246
K	1011	786	545	409	308	270	229	228
L	317	196	179	124	136	137	155	139
M	378	337	259	374	375	355	264	216
N	408	294	284	332	210	277	297	174
O	425	420	290	292	193	236	145	165
P	822	440	405	385	488	264	224	258
Q	642	364	302	217	224	194	162	129
R	606	232	205	209	187	183	205	166
S	423	357	440	322	295	310	351	190
T	282	224	285	120	230	89	122	75
U	457	525	423	479	462	379	280	262
V	420	345	208	245	177	224	184	105
W	133	111	87	159	103	101	167	53
Average	556	342	324	309	279.8	256.1	243.6	197.1

CHAPTER THREE

INTER-CORRELATIONS OF THE TESTS

In the course of the experimental investigation of general intelligence, a number of theories concerning the correlation of mental abilities have been advanced. Two opposing hypotheses may be defined—the compensation theory and the theory of the positive correlation of desirable traits. Or the problem may be observed from a slightly different angle, and one may distinguish three hypotheses, the theory of specific habits, of types or levels, and of the common factor.

The theory of compensation has, of course, been abandoned by psychologists, though it is still retained in such popular beliefs as that slow learning means long retention, or that ability at abstract thinking will incapacitate one for housework or gardening. Inverse relations are not, however, unknown. Such instances as Chapman's[1] negative correlation between cancellation and multiplication, and McCall's[2] negative correlation between cancellation and a number of other mental tests, may be advanced to show, at the will of the critic, either the probability of the existence of such inverse relations, or the unreliability of the experiments. In the majority of cases positive correlation seems to occur.

The hypothesis of the specification of traits has, like the compensation theory, been abandoned. It is no longer held that each mental function is a separate entity which may be developed independently, without effect upon other abilities.

Concerning the cause of such positive correlation as is observed, the two theories, the doctrine of levels and of the common factor are in dispute today. The difference seems to be main one of emphasis. Spearman,[3] who is the exponent of the "common factor," writes: "Here the view supported is that all performances depend to a certain degree upon one and the same general common factor, provisionally termed 'Gen-

[1]Chapman, J. C.—Individual Differences in Ability and Improvement and their Correlations. C. U. Contrib. to Ed.

[2]McCall.—Correlations of some Psychological and Educational Measurements—T. C. Contrib. to Ed.

[3]Hart and Spearman.—General Ability, Its Existence and Nature B. J. P. Volume 5.

eral Ability'. Correlations are thus produced between all sorts of performances, the amount of correlation being simply proportional to the extent that the performances concerned involve the use of this general common factor or 'General Ability.' The common factor is more closely related to each function than any trait is to any other trait." Spearman admits, however, special relating forces which, though subordinate in importance to the common factor, may also make for positive correlation.

Thorndike[2] emphasizes these special relating forces rather than the common factor. He says, speaking of the possibility of discovering the relation between traits. "Other things being equal two functions would be correlated in proportion as efficiency in them depends upon the status of the unknown 'Common Factor' variously called General Intellect, Mental Energy, Ability to Learn, and the like. Perhaps if we could define this factor, learn its symptoms and test its strength in different individuals, we should find that it alone explained the main features of all cross lines (representing relations.) Perhaps on the other hand, the other relating force would be even more important." One of the special relating forces is the organization of neurones into sensory and associative levels. "Correlations seem to be closer within the analytical or abstracting functions than between these and others. So also within the purely mental associative functions like adding, completing words, giving opposites and naming objects, than between one of them and one of the sensori-motor functions. The sensations seem to interrelate only loosely; and any one of them would relate very loosely to the associative or analytic functions, even when the latter was busied with data from that sense." Besides this organization into levels, other relating forces might make for positive correlation. Examples of these would be instinctive tendencies—as interest in persons might produce high correlations between love of ceremonies, ability in sociology and interest in literature—and the influence of special training—as the correlation between knowledge of Latin and of Geometry might be caused by the training of many individuals in both lines.

Almost every psychologist who has given more than two or three tests to a number of individuals, has made some contribution to the psychology of correlation. Those investigations

[2]Thorndike.—Educational Psychology, Volume III.

which have been most fruitful from the point of view of general theory have been attempts in the main to answer one or all of these three questions: 1st—Is there a positive correlation of desirable traits? 2nd—If such a positive correlation exists, is it due largely to the presence of a common factor or to special relating forces? 3rd—If a common factor does exist, what is its nature? Three kinds of experimental procedure have been adopted for the solution of these problems. The first method is that of giving a number of tests to a homogenous group of subjects and correlating the results of the tests with each other and with some estimate of intelligence. Ability in the second group of investigations is judged mainly by status as determined by achievement in life. There a notably good and a notably poor group are studied, and the main object of the investigation is to determine the difference in reaction of the two. The third type of study is not concerned with estimates of intelligence or with previously demonstrated ability, but simply with the relation between the traits measured by the tests themselves. Such investigations have frequently been undertaken for some other purpose—as to measure the effect of humidity or of caffein upon efficiency. Only as a secondary problem has the question of correlation been attacked.

Among the notable experiments of the first type are those of Burt, Wyatt, and Brown. Here the main interest seems to be in the definition of intelligence.

Burt* gave 12 tests to 43 boys. The tests measured sensory discrimination, motor ability, sensori-motor functions, association and voluntary attention. Teachers and students estimated the intelligence of the 43 boys. Burt correlated these estimates of intelligence with the records made by the boys in the various tests. The test of voluntary attention showed the highest correlation, association and memory tests came next, then sensori-motor, then motor, and finally tests of sensory discrimination showed the lowest correlations of all. He concludes from his results that all the functions of the human mind, the simplest and most complicated alike are processes within a single system. A process typical of higher psycho-physical levels may however, be connected with a process typical of lower psychophysical levels far less intimately than either is with a process of intermediate levels.

*Burt—Experimental Tests of General Intelligence—Br. Jr. Ps. 3.

In this Burt seems to favor the theory of levels but his emphasis—like Spearman's—is upon the common factor rather than upon the special relating forces. Voluntary attention, especially the power of readjustment to relatively novel situations is, for him, the essential factor in this general intelligence.

Wyatt[2] gave 15 tests to 75 children, giving each test twice. Analogies and completion tests, which seemed to measure functions of the higher levels, correlated most closely with estimated intelligence. Memory tests showed the next highest correlations. Wyatt, too, emphasizes the common factor.

Brown[3] who examined 259 children with 12 tests—each one given twice—found that first the completion test and next the mechanical memory tests, correlated most closely with general intelligence. Brown, unlike the other two, found no evidence of the existence of the common factor.

The second type of investigation includes such studies as those of Norsworthy, Terman, Binet and Simpson. Norsworthy,[4] in comparing defective with normal children, found that the order in which the different abilities tested would correlate with intelligence was: 1st—abstraction and association; then, memory, then various forms of perception; and last, motor control. Terman[5] tested seven bright and seven dull boys and found the bright boys superior in all the mental tests as puzzles, word-building, solving problems. Superiority of the dull boys in motor tests was explained as due to greater maturity, or different interests. Binet,[6] in testing five intelligent and six unintelligent pupils, found that tests of accuracy in tactile sensibility, counting rhythmic sounds, copying figures, sentences and drawings, memory of figures and cancellation of letters best differentiated the bright boys from the dull. He believed that quickness of adaptation was the important factor in intelligence.

Simpson,[7] made an intensive study of two groups of persons—a good group made up of 17 professors and advanced students of Columbia and a poor group of 20 men who had

[2]Wyatt—The Quantitative Investigation of Higher Mental Processes. B. J. P. 6.

[3]Brown—Some Experimental Results in the Correlation of Mental Abilities, B. J. P. 3.

[4]Norsworthy—The Psychology of Mentally Deficient Children.

[5]Terman—Genius and Stupidity—Pd. S. C. 13.

[6]Binet—"Attention et Adaptation" Annee Psychol. 6,248.

[7]Simpson—"Correlations of Mental Abilities" T. C. Contrib. to Ed.

never held any position demanding a high grade of intelligence, some from the Salvation Army Industrial Home, some from a Mission on the Bowery. He finds no hierachial arrangement of abilities, but does find a close relation among certain mental abilities which would imply a something which might be called "general intelligence." Certain capacities are, however, relatively specialized. The abilities which correlate most closely with other traits as shown by the fifteen tests given are—first, selective thinking, then memory and association, quickness and accuracy of perception, motor control, sensory discrimination. General intelligence, then, implies these different abilities in the order named.

The foregoing results (with the exception of Binet's) and a number of similar ones are based upon the scores of initial tests, upon records corrected for attenuation, or scores attained after a preliminary fore-exercise. Correlations vary from such coefficients as 94 (raw coefficients) obtained by Simpson between memory of words and Ebbinghaus test (when both groups were combined) to such numbers as 62 obtained by Burt between sorting and sound discrimination, and 25 obtained by Brown between accuracy in addition and the Ebbinghaus test. Are such figures a valid measure of the relation between the traits studied? Obviously they are not. The first trial is a poor measure of ability. Correction for attenuation is a statistical method for eliminating the influence of variability—it needs, however, confirmation from actual data. The method of giving a few practice trials does away with the influence of "adaptation." Various other conditioning factors (as previous experience of some individuals) are not excluded by this method.

Sufficient practice in the capacities to be observed might eliminate these sources of error. A score which was the result of a number of testings would be more reliable than a score based upon one or two tests. Taken at the close of the practice period, it would be subject to less variability. Since the amount of practice to be added is the same for each individual, the increment would serve to equalize the opportunity of each. In such an investigation no hypotheses concerning the nature of general intelligence or the cause of correlations observed to exist need be advanced. Such a study need throw light upon the problem of the existence of the common factor, only in so far as these further experiments give a

basis for questioning or confirming former results. When we ask what influence a further factor—as practice in a number of tests—has upon their intercorrelations, we are inquiring whether coefficients already found are reliable and representative figures.

Five practice experiments may be reported. A number of other studies as those of Whitley, Wells, Brown, Chapman, Binet, and Thorndike (see next chapter) bear only indirectly on the problem.

Spearman and Krueger[1] found that correlations for 8 successive quarter hours between such abilities as cancellation, adding, counting, reading, learning by heart tended to increase in size as practice increased, up to a certain point, after which they tended to diminish. The decrease might be due to fatigue entering as a disturbing factor. Where practice means improvement in individual scores correlations seem to increase. After fatigue enters, they diminish.

Burt[2] found that the second trial of eleven tests correlated less closely with estimated intelligence than did the first trial. As Jones points out this may be due to some factor of adaptation, common to all tests, which makes intercorrelations between tests at the first trial higher than would otherwise be the case.

Abelson[3] tested subnormal children and found the correlations between tests to be 32 (initial trial), 36, 37, and 40 at the second, third, and fourth trials.

Winch[4] studied the effect of practice on correlation between a simple motor test of cancelling all letters, and a complex motor test, cancelling a. n, o, s. The correlations on six successive days were 29, 44, 59, 48, 50, 47—an observable increase up to the third day.

Jones[5] in a study of eight tests made upon 203 boys, for four successive years, finds "no marked tendency for correlations of test measurements either to increase or to decrease on account of age and experience, under the conditions of the ex-

[1]Spearman and Krueger—Die Korrelation zwischen verschiedenen geistigen Leistungsfahigkeiten Z. f. P. 44.

[2]Burt—"Experimental tests of General Intelligence"—B. J. P. 3.

[3]Abelson—"The Measurement of Mental Ability of Backward Children, B. J. P. 4.

[4]Winch—Reported in Brown—"The Effect of Observational Errors" B. J. P. 6.

[5]Jones—"Influence of Age and Experience on Correlations."

periment. There are many striking irregularities which are difficult to explain in the course of the correlation from one year to the next."

"On the whole," he says, "it is evident that many tests decrease their correlations from the first to the second year, due presumably to the factor of the understanding of instructions (common to all tests at the time of their first administration). After this initial drop in the amount of correlation between the tests, there is likely to be a slight increase, in the following years, in the amount of correation. Whether this is due to practice alone or whether to the factors of age and vocational experience varying widely among the subjects, it is impossible to say. We are inclined to believe that the vocational life in Cincinnati during these three years does aid in differentiating the good from the bad. This is in spite of the fact that, in the case of certain types of tests, there is an evident evening up process (as in the memory tests) so that those who are proficient in one are not correspondingly proficient in the other, even after practice. The increase in the amount of correlation between different tests is greater when the tests are apparently unlike each other than when the tests are alike."

The results of Burt and Abelson are based each on only three or four trials. Spearman and Kreuger's results are complicated by the entrance of fatigue, Jones' correlations are influenced not only by practice but by the entire experience of the subjects during the year that intervened between testings. Hollingworth's* results are obtained from a very large number of trials. His experiment was such that it was possible to isolate the effect of practice from fatigue, growth, or general experience.

The following table was obtained by averaging the intercorrelations between six tests given to 13 individuals 205 times. After the initial trial, each score used is the average of five records.

TABLE NO. 9

Intercorrelations Between Tests in Hollingworth's Experiment.

	Adding	Opposites	Color Naming	Discrimination	Co-ordination	Tapping	Final Average
1	19	10	15	-07	-15	17	065
5	41	26	15	35	21	32	280
25	50	35	43	27	08	35	320
80	55	42	53	31	18	34	390
205	48	62	61	35	34	52	490

Correlations obviously increase. The change from first to second is, as Jones has pointed out, due partly to the fact that the coefficient for the first trial is based upon one score, for the second period upon an average of five scores. A decrease in variability—as these coefficients are uncorrected—may, similarly, be a factor in the increase. Hollingworth* suggests as causes in addition to this change in variability a change in the nature of the tests themselves, and possibly some explanation in terms of a common fund of energy or at least of the existence, when final ability is approximated, of a positive relation between desirable traits.

It is the object of the present chapter to investigate the effect of practice somewhat more minutely, to confirm or question the former results, and to suggest, if possible, some cause for whatever relation is found to exist.

In order to eliminate the change in the nature of the tests, a factor which seems to be of some importance in Hollingworth's results, functions were chosen which, it is believed, did not become more similar as repetitions were increased. Hollingworth's opposites and calculation test, for example, came, after identical repetition to resemble greatly the color-naming test. No conscious choosing of an opposite or mathematical operation was involved. Each stimulus on the card was bound to one definite response. With the exception of the larger number of stimulus-response series involved, and possible inequality of previous practice, these tests gradually approached in character the color-naming test. Two of the tests of the present investigation, as was noted above, cannot become automatic through repetition. The multiplying and word-building tests involve a response each time, to a new situation. Nor do these functions come to resemble each other or the color-naming, tapping, and adding tests as time goes on. Only the adding and color-naming tests might seem to approach each other. Part of the correlation between color-naming and adding might possibly be due to increased similarity. But the amount of initial likeness is not great—not nearly so important as in the tests in the former investigation—and the amount of practice given is probably insufficient for a reduction of the adding operation to a practically automatic response.

*Hollingworth—"Correlation of Abilities as affected by Practice." Jr. Ed. Psy. 4.

The practice series was divided into eight periods and each test was correlated with every other test at each of the eight divisions. Table 10 gives the result of this correlation. There the median record (in all but the first trial) is the score used.

TABLE NO. 10

Intercorrelations Between Tests (Median the Measure). Average along side without multiplying. Average at bottom with multiplying.

	C. N	Tap.	Add.	Mul.	W. B.	Av.
1st Division						
C. N.	x	32	20	–02	21	24
Tap.	32	x	24	–16	12	23
Add.	20	24	x	29	24	23
Mul.	–02	–16	29	x	21	
W. B.	21	12	24	20	x	19
Av.	19	13	24	08	19	
2nd Division						
C. N.	x	32	32	–12	21	28
Tap.	32	x	–21	–02	20	10
Add.	32	–21	x	51	42	18
Mul.	–12	–02	51	x	32	
W. B.	21	20	42	32	x	29
Av.	18	07	26	17	28	
3rd Division						
C. N.	x	36	35	03	08	26
Tap.	36	x	24	–16	05	22
Add.	35	24	x	48	27	29
Mul.	03	–16	48	x	26	x
W. B.	08	05	27	26	x	13
Av.	21	12	34	15	17	
4th Division						
C. N.	x	33	41	17	35	36
Tap.	33	x	–01	–05	25	19
Add.	41	–01	x	78	33	24
Mul.	17	–05	78	x	17	
W. B.	35	25	33	17	x	31
Av.	32	13	38	27	28	
5th Division						
C. N.	x	63	35	–04	33	44
Tap.	63	x	26	10	40	43
Add.	35	26	x	76	46	36
Mul.	–04	10	76	x	07	
W. B.	33	40	46	07	x	40
Av.	32	35	46	22	32	
6th Division						
C. N.	x	67	44	11	22	44
Tap.	67	x	25	05	30	40
Add.	44	25	x	56	33	34
Mul.	11	05	56	x	28	
W. B.	22	29	33	28	x	28
Av.	36	32	40	25	28	
7th Division						
C. N.	x	42	37	16	41	40
Tap.	42	x	36	–02	03	27
Add.	37	36	x	47	20	31
Mul.	16	–02	47	x	13	
W. B.	41	03	20	13	x	21
Av.	34	20	35	19	19	

8th Division

C. N.	x	46	36	08	16	33
Tap.	46	x	22	09	–07	20
Add.	36	22	x	62	39	32
Mul.	08	09	62	x	17	
W. B.	16	–07	39	17	x	16
Av.	27	18	40	24	16	

Table 11 gives the average correlation of each test with every other test at each of the eight divisions.

TABLE NO. 11

Average Correlation of Each Test with Every Other Test.

	1	2	3	4	5	6	7	8
C. N.	19	18	21	32	32	36	34	27
Tap.	12	01	12	13	35	32	20	18
Add.	24	26	34	38	46	40	35	40
Mul.	08	17	15	27	22	25	19	24
W. B.	19	28	17	28	32	28	19	16
Av.	16	19	20	28	33	32	26	25
Ave. of Two.			24		33		26	

There seems to be a gradual increase in the average of all coefficients which lasts up to the sixth period, after which there is a decrease. The averages for the separate tests show this. Color-naming and multiplying increase up to the sixth, tapping, adding and word-building up to the fifth. An observation of the original coefficients in Table 4 confirms this. The correlation between tapping and adding reaches its maximum at the seventh division, between tapping and color-naming, color-naming and adding, word-building and multiplying at the sixth, between tapping and multiplying, tapping and word-building, word-building and adding at the fifth, between color-naming and multiplying, mutiplying and adding at the fourth, between word-building and multiplying at the second. After the maximum point there is a slight decrease in correlations.

The multiplying test has a low and occasionally negative correlation with all tests except adding. The unreliability of the test as compared with others (see chapter 3) may account for some of this observed lack of relation. Another table from which all correlations with multiplying were omited in averaging is given here.

With the exception of adding (and occasionally of word-building) all average coefficients are higher when multiplying is omitted. These correlations increase as did the others, up to the fifth or sixth division, after which there is again a slight decrease.

TABLE NO. 12

Average Coefficients—Multiplying Omitted.

Test	1	2	3	4	5	6	7	8
C. N.	24	28	26	36	44	44	40	38
Tap.	23	10	22	19	43	40	27	20
Add.	23	18	29	24	36	34	31	32
W. B.	19	29	13	31	40	28	21	16
Ave.	22	21	23	28	41	37	30	25
Average of Two			26		39		28	

Beside this median record another measure of individual achievement was employed. Best scores instead of median records in each practice period were correlated. Such scores are (See chapter 3) no more variable than median records, and this in spite of the fact that they are determinations based upon only one testing. They would seem at first sight to have the advantage of being most representative of ultimate achievement. A tendency to inaccuracy would likewise be immediately suggested, though such a suggestion seems to be contradicted by the relatively high adjacent correlation. Is there any difference in the size of correlations obtained by this method? What factors operate to produce such relations as are observed?

Table 13 gives the result of inter-correlations of best records in the eight periods.

TABLE NO. 13

Intercorrelations Between Tests (Best Record Used as Measure). Average without multiplying at side, with multiplying at bottom.

	C. N.	Tap.	Add.	Mul.	W. B.	Av.
			2nd Division			
C. N.	x	31	34	–05	15	27
Tap.	31	x	–10	01	22	14..
Add.	34	–10	x	46	37	20
Mul.	–05	01	46	x	16	
W. B.	15	22	37	16	x	25
Av.	19	11	27	15	23	
			3rd Division			
C. N.	x	28	50	38	35	38
Tap.	28	x	06	–20	28	21
Add.	50	06	x	67	11	22
Mul.	38	–20	67	x	–70	
W. B.	35	28	11	–70	x	25
Av.	38	11	34	04	01	
			4th Division			
C. N.	x	43	48	13	23	38
Tap.	43	x	07	–01	17	22
Add.	48	07	x	60	40	32
Mul.	13	–01	60	x	05	
W. B.	23	17	40	05	x	27
Av.	32	17	39	19	21	

			5th Division			
C. N.	x	57	45	04	43	48
Tap.	57	x	19	06	37	38
Add.	45	19	x	65	53	39
Mul.	04	05	65	x	03	
W. B.	43	37	53	03	x	44
Av.	37	30	46	20	34	
			6th Division			
C. N.	x	57	47	–14	30	45
Tap.	57	x	33	–19	20	37
Add.	47	33	x	58	35	38
Mul.	–14	–19	58	x	21	
W. B.	30	20	35	21	x	28
Av.	30	23	43	12	27	
			7th Division			
C. N.	x	62	42	–04	34	46
Tap.	62	x	25	–23	–04	28
Add.	42	25	x	18	28	32
Mul.	–03	–23	19	x	07	
W. B.	34	–04	28	07	x	19
Av.	34	15	29	–01	16	
			8th Division			
C. N.	x	50	44	14	30	41
Tap.	50	x	26	–20	39	38
Add.	44	26	x	52	36	35
Mul.	14	–20	52	x	–01	
W. B.	30	39	36	–01	x	35
Av.	35	24	40	11	26	

In Table 14 the correlations of each test with every other test are averaged. In Table 15 the averages are given without multiplying.

TABLE NO. 14

Average Correlation of Each Test with Every Other Test (Best Record used as Measure).

Test	2	3	4	5	6	7	8
C. N.	19	38	32	37	30	34	35
Tap.	11	11	17	30	23	15	24
Add.	27	34	39	46	43	29	40
Mul.	15	04	19	20	12	–01	11
W. B.	23	01	21	34	27	16	26
Av.	19	18	28	33	27	19	27
Av. of Two		23		30		23	

TABLE NO. 15

Average Correlation of Each Test, with Every Other Test, excluding Multiplying (Best Record used as Measure).

Test	2	3	4	5	6	7	8
C. N.	27	38	38	48	45	46	41
Tap.	14	21	22	38	37	28	38
Add.	20	22	32	39	38	32	35
W. B.	25	25	27	44	29	19	35
Av.	22	27	30	42	37	31	37
Av. of Two		29		40		34	

A comparison of these tables with those of median records shows them to be similar, though some correlations by the best record method are higher, some lower. The average coefficients are a little lower by the best record method when multiplying is included, a little higher when multiplying is omitted. This corresponds to the result that correlations by the best records method are less reliable in multiplying than by the median method. Correlations with word-building are similarly occasionally lower and are (see table 3) less reliable. Correlations during the eighth practice period are a little higher for the "best" series. For three tests the reliability is greater. "Best" records show, then, neither a higher nor a lower correlation than median records. The same progression is observable—an increase up to the fifth or sixth periods, followed by a slight decrease.

Still another measure may be used. We may do away with the eight divisions of the practice period and consider each subject's highest point only. The best score of each individual in each test, whether this record is attained in the beginning, middle, or at the end of the practice series has been correlated with her best record in every other test. The variability of such a record cannot be determined. Presumably it is large, since such a score would be subject to many possible errors. Accidental factors would lower the correlation below its true amount. The table follows:

TABLE NO. 16

Correlation of each Test with every other at Highest Point reached by each subject.

	C. N.	Tap.	Add.	Mul.	W. B.
C. N.		52	52	05	27
Tap.	52		33	–19	23
Add.	52	33		36	32
Mul.	05	–19	36		01
W. B.	27	23	32	01	
Av.	34	22	38	06	21
Av. (without M.)	44	36	39		27

The average of correlations with multiplying is .24, without multiplying it is .37. The .24 is equivalent to the average for the correlation of medians of the third and fourth division, the .37 is nearest the .39 obtained by averaging correlations in the fifth and sixth divisions, when multiplying is omitted. Uncorrected correlations of best records show, with the exception of multiplying, a relatively high correlation,

about equal to that obtained by the median record method at its highest point.

The cause of the increase and decrease of correlations, observed both in median and best records, of the equal reliability of median and best records, of the relatively high correlation of very best records when compared with their probably unreliability, is difficult to find. Rise and fall in coefficients have been attributed to improvement through practice, which increases correlations, up to the time of the entrance of another factor, fatigue. Jones believed that a decline in correlation was due to the elimination of the adaptation factor which raised the initial correlation above its true amount. Hollingworth's results showed a steady increase in coefficients which he attributed partly to lessened variability, partly to change in the nature of the tests, and partly to the fact that final tests were more representative of ultimate capacity than were initial trials.

In the present results a parallelism between the change in size of intercorrelations, the change in size of adjacent correlations, and the amount of gain made by practice is at once apparent. Up to the fifth or sixth trials individual and average scores (see Chapter 3) show a great improvement. After this there is an improvement level, or in some cases retrogression. Adjacent correlations are lowest at the beginning of practice, higher at the very end, but greatest of all at the fifth or sixth practice period. Similarly intercorrelations show an increase and subsequent slight decline.

The coincidence of the changes would tend to suggest one progression as the cause of the others. Obviously the increase in individual scores is the primary factor. Without change in records there can be no variation in correlations based on these records. While great progress is being made, variability is great and intercorrelations low, at the time of maximum achievement variability is least and intercorrelation highest, where achievement is somewhat reduced there appears a somewhat greater variability and a somewhat lower intercorrelation.

A change in variability may be suggested as the entire cause of the change in intercorrelation. If this is the case, then when correlations are corrected for attenuation, the differences should disappear and beginning, end, and middle correlations should be approximately equal. Table 14 gives

the corrected coefficients for the second, fifth and eighth periods.

TABLE NO. 17

Corrected Coefficients (Median) 2nd Period.

	C. N.	Tap.	Add.	Mul.	W. B.
C. N.	x	40	36	–16	25
Tap.	40	x	–27	–03	27
Add.	36	–27	x	69	51
Mul.	–16	–03	69	x	44
W. B.	25	27	51	44	x
Av.	24.6				
Av. without M.	28				

5th Period

	C. N.	Tap.	Add.	Mul.	W. B.
C. N.	x	69	37	–05	39
Tap.	69	x	27	12	48
Add.	37	27	x	93	55
Mul.	–05	12	93	x	10
W. B.	39	48	55	10	x
Av.	38.5				
Av. without M.	46.0				

8th Period

C. N.	x	51	40	09	18
Tap.	51	x	27	11	–08
Add.	40	27	x	78	47
Mul.	09	11	78	x	21
W. B.	18	08	47	21	x
Av.		29.4			
Av. without M.	29.2				

Coefficients still vary. The average with multiplying progresses from 25 to 39 to 30, without multiplying from 28 to 46 to 29. This is to be compared with the series 19, 33 and 25, and 21, 41 and 25 of the uncorrected medians.

	Δ2 & 5	Δ5 & 8
Median Uncorrected (with M)	14	–8
Median Corrected (with M)	14	–9
Median Uncorrected (without M)	20	–16
Median Corrected (without M)	18	–17

The difference then, does not disappear. Coefficients are on the average, raised but the gross difference between the averages of the correlations of the different periods remains unchanged. The increase in inter-correlation cannot, then, be attributed wholly to a change in variability.

Other suggestions pointing to the activity of some other factor, would be the equal reliability of best as compared with median records, and the somewhat superior inter-correlations obtained by working with very best scores.

We must suppose that approach to an improvement level gives not only decreased variability but equalizes the effect of previous practice and enables the individual's ultimate capacity to appear more plainly. We might say that best scores in each period are no more variable than determinations based upon more than one score, just because the former are more representative of final ability. Similarly very best records which seem to be the best evidence of what ultimate achievement might be, show, even though uncorrected, high inter-correlation. Not accuracy alone but the quality of representing ultimate capacity, seems to be the characteristic of records which produce higher inter-correlations.

In this connection one should note the meaning of the coefficient of reliability which has been used. It is a measure of the change in relative position from one median record to the next. Where change is greatest, as at the beginning of practice this coefficient will be lower, not only because records vary more but because some individuals are gaining more, some less. Where there is relative stability the coefficient is high, where some retrogression it is lower again. The use of it for correction would seem to raise initial correlations above their true amount, for not only variability but different amounts of improvement are excluded. At the end of practice a high adjacent coefficient is caused not only by less variability but by greater equality of gain from practice.

It may be, then, that correlations tend to increase as final ability is approximated. Final capacity probably best represents true capacity, ability freed from such influences as accidental initial high or low performance, variability of a few trials, or inequality of previous practice. True measures of individuals would presumably show a positive and fairly high correlation. A number of the negative correlations found may be due to the inaccuracy of the measures. But again, it is important to remember that these correlations are not, except in a few cases, remarkably high. Occasionally the change is less than the probable error of the median coefficient. That chance is not entirely responsible for the change is evidenced by the constancy of direction of the variation. That the correlations are as high as they are, in view of the relative homogeniety of the subjects is a fact pointing toward a possible finding of great positive relationship between abilities tested in a random sampling of humanity.

An examination of individual coefficients should be added to determine if possible, the functions operative in the increase and to suggest as well as so few tests can, the different degrees of relation between traits. The median uncorrected table will be used throughout.

The correlation between color-naming and tapping is always positive and high. It increases up to 67 in the sixth division. Color-naming and adding begin with a correlation of 20 which increases up to 41. Color-naming and multiplying begin with a negative correlation which remains less than twenty throughout the series. The correlation between color-naming and word-building begins at 21, goes down to 08, but reaches finally 41. Tapping and adding begin at 24, correlate twice negatively, finally reach 36. Tapping and multiplying begin at –16, remain minus till the fifth trial where they become slightly positive, never becoming, however, greater than 10. Tapping and word-building begin at 12, become negative, finally reach 40 after which they are again negative. Adding and multiplying always have a high positive correlation which begins at 30 and increases to 78. Addng and word-building change from 24 to 46. Word-building and multiplying increase from 20 to 32 but remain irregular throughout.

Results from so few tests are of little value for the solution of general theoretical problems concerning the relation between traits. A few facts may be observed.

The importance of identical elements and of content is notable in the high correlation of adding and multiplying. In spite of the unrelability of the multiplying test, uncorrected coefficients of adding and multiplying are high—due probably both to native ability and special training in a definite content, and to the identity of the elemental operations. A low correlation of multiplying with color-naming and tapping is observed. Is multiplying possibly an ability of a higher level, and is the lack of correlation due to difference in the kind of performance? Multiplying and word-building, both of which are tests of the analytic and selective variety show a somewhat higher correlation. Perhaps the correlation would have been higher still had multiplying been aways a test of analysis, had the determination been more reliable, and the amount of practice greater. We have a very high correlation between color-naming and tapping—functions of the associative and sensori-motor level.

The averages with multiplying show that adding has the highest correlation with all other tests, color-naming the next, then word-building, multiplying, and tapping. When correlations with multiplying are omitted the order of greatest correlation is color-naming, adding, tapping, and word-building. We might attribute the high correlation of color-naming and adding to the relative lack of variabilty of the tests and to the fact that in these an improvement level was approximated.

We can make no conclusion as to general theory of the cause of the observed relationship. Correlations do seem to depend upon identity of elements, they are higher in the case of multiplying for a test of the same, than for a test of other levels. They are influenced by the amount of variability of the test, and, apparently, by approximation to proficiency.

SUMMARY

The uncorrected coefficients obtained in the present investigation confirm the suggestion offered by Hollingworth that practice increases the amount of inter-correlation.

The increase in size of coefficients coincides with increased individual and average scores which are the result of practice, and with decreased variability.

When coefficients are corrected for attenuation the difference does not disappear.

This increase may then, be attributed, to the fact that true measures of ultimate capacity are being approximated.

This is borne out by the fact that a measure which best represents ultimate capacity shows, in spite of its being probably subject to incidental variation, a relatively high correlation.

Those tests in which an improvement level is approximated and which are simpler and more reliable have on the whole, higher inter-correlations.

On the whole, tests of the same level correlate a little closer than do tests of different levels.

CHAPTER FOUR

THE RELATION OF IMPROVEMENT TO INITIAL ABILITY

If it were true that correlations between tests increase with approach to an improvement level, then we might say that practice with improvement tends in one sense, at least, to increase or to emphasize indvidual differences. Instead of finding one person capable in one direction, one in another, we might discover that some people were good in most lines, others poor in most lines, others mediocre. Taking the whole range of abilities into consideration it might be shown that the average status of each individual, when the compensating influence of special training was equalized, was more unlike the average status of each other individual after, than before practice gain.

We might, on the other hand, observe one instead of a number of functions, noting the influence of practice in that one ability upon differences as displayed in that trait. One way of approaching this problem would be to inquire—who improves most, the initially poor or the initially good individual? If the one who is most efficient in the beginning, makes the greatest gain, then practice would seem to increase differences. If the one who is originally poor, improves the most, then persons would seem to approach each other, as repetitions are increased. A finding of the first sort would seem to show that native ability evidenced, perhaps, by the first record made, rather than environment, in the form of special training, was the factor responsible for proficiency. A result of the second sort would seem to mean that, since the equalization of practice would decrease differences, inequality of training was the cause making for variation.

A number of studies have been made which deal more or less directly with this problem. Among those which bear definitely on the question, the investigations of Thorndike, Whitley, Wells, Brown, Kirby, Donovan, Hahn, Chapman and Dallenbach are of greatest interest.

Thorndike,* in 1908, investigated the effect of practice in the mental multiplication of one three-place by another three-

*Thorndike—The Effect of Practice in the Case of a Purely Intellectual Function, A. J. P. Vol. 19.

place number. Twenty-eight individuals solved ninety-six examples each. The subjects kept their own time records and errors were afterwards transmuted into time. Equality of practice, therefore, means an equal amount of work performed. In this function the physiological limit is far below the ability reached by any individual although, as Thorndike states, two persons seem to reach a plateau. The table showing the improvement made by individuals differing in initial ability is repeated here.

TABLE NO. 18

The Effect of Equal Amounts of Practice upon Individual Differences in the Mental Multiplication of a Three-place by a Three-place number.

		Amount done per Unit of Time.			Percentage of Correct Figures in Answer.		
	Hours of Practice	1st 5 Examples	Last 5 or 10 Examples	Gain	1st 5 Examples	Last 5 or 10 Examples	Gain
Initial highest 5 Individuals	5.1	85	147	61	70	78	18
Initial next 5 Individuals	5.1	56	107	51	68	78	10
Initial next 6 Individuals	5.3	46	68	22	74	82	8
Initial next 6 Individuals	5.4	38	46	8	58	70	12
Initial next 5 Individuals	5.2	31	57	26	47	67	20
Initial next 1 Individual	5.2	19	32	13	100	82	—18

The table shows that the best at the start made a larger gain in amount done per unit of time and an approximately equal gain in accuracy, with the exception of one individual.

A study of practice in adding seemed at first to bring opposite results.* The correlation between initial ability and percentile improvement was a negative one, $-\frac{1}{4}$, and for gross improvement (by my computation) –68. The ratio of highest to lowest individuals as given by Thorndike in the original article changes as shown in this table.

TABLE NO. 19

	Ratio of Highest to Lowest Individual				
	1st 19th	2nd 18th	3rd 17th	4th 16th	5th 15th
In first practice period	2.49	2.25	1.92	1.82	1.46
In second " "	3.36	1.99	1.87	1.62	1.46

*Thorndike.—Practice in the Case of Addition A. J. P. Vol. 2, pp. 488-486.

Thorndike says, "The differences amongst individuals in the ability to add seem to be due in larger measure to differences in environmental influences: For equal practice does reduce a little the relative or percentile differences within our group." And farther on "There is, of course, no essential conflict between this result for addition and the opposite result for mental multiplication with two three place numbers. The same theoretical view which would expect mental span and ability to manage very complex relationship in a given field to be increased by practice in close dependence upon original capacity would expect particular associative habits such as thinking of thirteen upon seeing 4, 7 and 2 in a column to be increased by practice in less close dependence upon original capacity."

These results were obtained and computations made in terms of amount of time required to perform a definite task. In 1914 the results are again reported.* There the average number of additions per 5 minutes, corrected for errors, is the score used and each of the nineteen subjects is given an approximately equal time for practice. The table follows.

TABLE NO. 20

The Effect of Equal Amounts of Practice upon Individual Differences in Column-Additions of One-Place Numbers.

	Average Number of Additions per 5 minutes corrected for errors			Average time spent in practice from mid-point of first test to mid-point of last test (in minutes)
	First Test	Last Test	Gain	
Initially highest 6 Individuals	297	437	140	40
Initially next highest 6 Individuals	234	345	111	49
Initially lowest 7 Individuals	167—	220+	54	46

As the facts are here presented, it is apparent that the initially best gain the most.

When, then, decreased variability after practice and lack of correlation between initial ability and improvability computed in terms of time required are used as measures of the effect of practice, individual differences seem to decrease with repetition. When the subjects are divided into groups on the basis of their initial ability, when the amount measure is used and the time of practice made approximately equal, the average improvement made by each of these groups seems to show that equal amounts of practice are effective in increasing individual differences.

*Thorndike—Educational Psychology, Vol. 3.

The results of Whitley's[1] correlations are in substantial agreement with those of Thorndike. After an analysis of methods of measuring changes in which she shows the different results which may be obtained by measuring gain in time, in amount, in per cent, and by two variations, (the Smythe Johnson and Amberg) of the percentile method, she gives results obtained from five tests practiced by nine women for twenty days. The tests were the familiar curved maze, mental multiplication of two three place numbers, a sorting test, a cancellation test and a lifted weight test. Correlations between position at start and at the end of practice, and position at the start and average position, or gross gain or percentile gain, were:

TABLE NO. 21

	Pos. at start & at Finish	Pos. at start & Av. Pos.	Pos. at start & Gross Gain	Pos. at start & Per. Gain
Lifted Weights	.27	.45	−.98	−.65
Cancellation Tests	.72	.58	−.90	−.38
Sorting Test	.56	.58	−.92	−.36
Mul. Test	.44	.58	−.63	−.52
Maze Test	−.21	.33	−.95	−.90

There is a low positive correlation between initial and final ability in all these tests but the maze and a slightly higher correlation between position at the start and average position. In all tests there is a relatively high negative correlation between gain and rank in the first trial. Whitley says of these results: "Individuals with low standing can and do improve the most judging objectively though even so, they may not in conveniently measurable periods of time overtake those whose standing was high at the beginning."

In concluding her discussion of the improvement curve she says: "All functions do approximate one typical law for changes in improvement though individuals may tend to differ in variability. From this point of view practice must tend to make people more alike."

Thorndike in 1914 says of Whitley's results:[2] "The same effect (meaning an increase of difference due to practice) appears though less emphatically in the case of Whitley's nine individuals in a similar experiment. The four who were most efficient at the start made a greater average gain from equal

[1]Whitley—An Empirical Study of Certain Tests for Individual Differences, Archives of Psychology.

[2]Thorndike.—Education Psychology.—Volume 3.

practice than the four who were least efficient." Again averaging the records of groups of persons gives results apparently contradictory to those obtained by the correlation method.

Wells,[1] (1912), made a study of the effect of repetition in an adding and a cancellation test, using five men and five women as subjects. His results, based on a comparison of the improvement curves obtained, show that when absolute gain is considered, both men and women shows an association of high initial ability with possibility of great improvement. It is notable, however, that all of these comparisons are subject to exceptions—an individual whose record is good at first may make little gain, or an individual whose original ability is poor may make a surprisingly large gain from practice. Measures by relative gains show "no negative relation" with gross performance. The Table which Wells gives showing the orders obtained by the different measurements is given here.

TABLE NO. 22

Well's Table (Individuals designated by Names of Colors)

Initial Performance		Absolute Gain		Relative Gain	
Addition Test (Men)					
Blue	1	Blue	1	Red	4
Brown	2	Brown	2	Green	5
Orange	3	Red	4	Brown	2
Red	4	Green	5	Blue	1
Green	5	Orange	3	Orange	3
Additional Test (Women)					
Blue	1	Red	3	Red	3
Green	2	Blue	1	Blue	1
Red	3	Green	2	Brown	5
Orange	4	Orange	4	Orange	4
Brown	5	Brown	5	Green	2
Number-Checking Test (Men)					
Green	1	Brown	2	Brown	2
Brown	2	Green	1	Red	4
Blue	3	Red	4	Green	1
Red	4	Blue	3	Orange	5
Orange	5	Orange	5	Blue	3
Number-Checking Test (Women)					
Red	1	Blue	4	Brown	5
Green	2	Orange	3	Orange	3
Orange	3	Green	2	Blue	4
Blue	4	Brown	5	Green	2
Brown	5	Red	1	Red	1

[1]Wells.—The Relation of Practice to Individual Differences, A. J. P. Vol. 23, pp. 75-88.

The figures whch I have added giving the initial rank of the individuals at the right of the columns shows the changes even more clearly. Red (Number-Checking-women) changes from first to fifth position when initial ability and gross per cent gain are compared. Orange and Brown (Number-Checking Men & Addition Women), on the other hand, are fifth in two cases. The table shows both negative and positive results.

Wells adds to his comparison of improvement curves, a discussion of the value of different measures of improvability, favoring absolute measures. Factors in improvement seem to be, in the order of their influence in the present tests, (1) a difference in the individuals (resp. functions) plasticity, i. e., ability to profit by practice, (2) a difference in the actual amount of practice experienced and (3) constitutional factors, independent of plasticity, in the nervous system. He concludes:

"The more significant results would seem to be the indication that a superior performance at the beginning of special practice is not necessarily nor even probably attained at the sacrifice of prospects for future improvement. A high initial efficiency may carry with it as much as or more prospect of improvement under special practice as a low one. It was not because the favored individual had had more of the general experience enabling him to meet the experimental situation better but because he posssed the native abiity to profit more by such experience, general and special, past and future. Not practice, but practicility, is responsible for the superior position of such an individual and, in broader aspect, not education but educability."

Warner Brown* who studied improvement in card sorting, had twenty-six individuals make four or eight trials a day for thirteen days. In addition to results obtained concerning interference of habits he reports data on the relation between initial and final speed and improvability. His records showed (1) a high positive correlation between speed on the first day and speed on any succeeding day (2) a negative correlation between speed on the first day and ability to profit by practice, (3) a negative correlation between speed at the end of thirteen days and improvement made during this time." There are a

*Brown.—Interference in Card Sorting. University of California Publications.

few exceptions to the second finding—a few persons were both quick at the beginning and able to gain a great deal from practice. "Apart from these exceptional cases," Brown says, "it seems on the whole unquestionable that those persons who begin slowly will generally manifest a greater capacity for improvement (relative as well as absolute) than those persons who begin work at a high rate of speed. The effect of practice in this complex process is to flatten out the differences between individuals." On the other hand, he says: "Large practice gains do not insure the attainment of superior skill, on the contrary they serve as a rule only to reduce the handicap of those persons who are slow at the beginning."

Hahn and Thorndike[1] give the following table for amount of improvement made by school children in arithmetic. The numbers in the table refer to number of examples correctly solved. The numbers in parentheses designate the number of individuals considered.

TABLE NO. 23

Amount of Improvement in Relation to Initial Ability.

Initial Ability	Grade 4	5	6	7	Average of Grades
0-6	12.3 (10)	10.7 (7)	9.0 (29)		10.7
7-12	7.6 (9)	9.7 (12)	10.1 (9)	15.0 (2)	10.6
13-18	13.4 (5)	7.6 (8)	12.3 (6)	17.5 (2)	12.7
19-24	6.7 (3)	19.5 (4)	11.0 (6)	20.4 (9)	12.3
25-30		12.3 (4)	1.5 (2)	23.4 (6)	12.4
31-36		54.0 (1)	8.5 (2)	16.4 (5)	
37-42				9. (1)	

The authors conclude from this table that individual difference increase with practice since the initially more proficient gain on the average, more.

The variability of the averages obtained, is, however, very great. "Initial ability 25-30," varies from 1.5 to 23.4, "19-24" from 6.7 to 20.4, "31-36" from 8.5 to 54.0.

Data obtained by Donovan and Thorndike (1913)[2] gave similar results. The average gross gain in number of examples done for the four initially least proficient of 29 fourth grade pupils was 5.3, for the next four, 10, the next four, 5.3, the next seven, 6.6, next four, 8.3, next five, 6.8, next two, 11.

The authors say: "These results, showing so little power of equal addition to training to reduce individual differences, make it improbable that a very large function of the difference

[1]Hahn and Thorndike.—Some Results of Practice in Addition Under School Conditions. Journal of Education Psychology, Vol. 5 No. 2.

[2]Donovan & Thorndike.—Improvement in a Practice Experiment Under School Conditions, A. J. P. Vol. 24, pp. 426-428.

found among school children can be greatly attributed to difference in amount of training." One does not find, in this table, however, a definite tendency for the best persons at the start to improve more. The most that we can safely say is that there is an equal improvement of best and poorest.

Kirby's curves,[1] drawn on the basis of the performance of about 700 school chidren practising addition and division for equal amounts of time show a positive relation between gain and initial ability.

Chapman's[2] results for 22 individuals, five tests, show a negative correlation for the simpler functions which are nearer their physiological limit, a positive relation for the more complex functions.

TABLE NO. 24

Correlation between Initial Scores and Improvement.

	Initial Score	(1 2 3)
Column I	Improvement	(6 7 8) – (1 2 3)
Column II	Improvement	(9 1 0) – (1 2 3)

	Col. I Correlation	Col II Correlation
Color	03	–.14
Cancellation 2	06	05
Opposites	–.09	–.19
Addition	.44	.34
Multiplication	28	.41

The correlations between initial and final position are—color 87, cancellation (2) 75, cancellation (3) 85, opposites 50, addition 96, multiplication 87—all it should be observed, are high and positive. Chapman concludes: "It appears that in complex functions an individual who has gained high efficiency by previous general training will also improve correspondingly during the practice period, whereas in simpler functons as usually scored, initial efficiency probably bears little relation to improvability. If the method of scoring were altered so as to weigh improvement which is made as the subject approaches the physiological limit, there is little doubt that high correlations would be obtained in the simpler functions."

One recent study of normal and backward school children gave important results. There the function studied (Dallen-

[1]Kirby.—Practice in the case of School Children, Teachers College, Columbia University Contributions to Education No. 58.

[2]Chapman.—Individual Differences in Ability and Improvement, and their Correlations.—Col. U. Contributions to Education.

bach 1919)* is visual apprehension. Both normal and backward children were divided into three groups on the basis of their performance in initial tests. The amount of improvement computed for each group is shown in the following table.

TABLE NO. 25

Normal Group		Feeble-Minded Group	
Initial Ability	Improvement	Initial Ability	Improvement
42.72	13.74	48.5	6.12
35.32	13.43	41.26	3.14
28.56	15.71	34.31	2.31

The feeble-minded group were initially more proficient. They improved less, however, so that the normal group finally surpassed them. Dallenbach says: "The feeble-minded and poorer third of normal group continue to improve whereas the superior and mediocre subjects of the normal group show a large initial increase and then a long plateau." It seems improbable that a function like this should be near its physiological limit, yet when abnormal are compared with normal, or initially poorer normal with initially better normal—the individuals best at the start are seen to improve least. When comparisons are made within the feeble-minded group the opposite effect is observed.

In this problem then, results and especially conclusions drawn from them differ widely. The apparent contradictions should be due, it seems, to the fact that the expermientors studied different functions, permitted different amounts of repetition, used different methods of measuring gain and equating practice. Occasionally different conclusions may be drawn from the same experiment. A correlation figure may be urged in support of one assumption while an observation of the trend of average results may lead to another conclusion. We find, however, a certain amount of agreement. With the exception of Whitley's results, experiments on mental multiplication, adding, dividing, show in general, a positive, or, at the most, a zero relation between initial ability and gain. Improvement in tests like color-naming, opposites, cancellation, card-sorting, visual apprehension seems normally to be negatively related to initial proficiency. Those performances in which practice is measured in terms of equal time spent (as in Kirby's, Donovan and Thorndike's, Hahn and Thorndike's experiments) show more frequently a positive correlation than

*Dallenbach—A Test for Visual Apprehension Jour. Ed. Psy., 1919.

do those in which equality of practice means the reaching of a certain result a definite number of times (as in all color-naming tests, Whitley's and Brown's investigations and Chapman's opposites.) Exceptions to this are Dallenbach's visual apprehension and Chapman's cancellation test. In both cases, though equal practice meant equal time spent, the correlation was negative. Conclusions seem to depend upon the nature of the function studied and the measure of equality of practice used. The problem of the measure of improvement is similarly of importance.

As Thorndike, Whitley, Wells and others have pointed out, the question of the measurement of gain is a complex one. Theoretical difficulties are involved, as the problem as to whether gains made at different parts of the practice curve may be called equivalent. Even if we confine ourselves to actual time, amount, or percentile figures, asking not about relations of gain considered in the light of their frequency, difficulty, etc., but about gains compared on the basis of some numerical scores, questions of the interval over which improvement shall be measured and of the score which shall be used, remain. It might be well instead of choosing one measure of gain and basing all conclusions on that, to proceed empirically, to try out a number of methods and to compare the results obtained. Whitley has demonstrated the different conclusions reached when theoretical scores are treated by different methods of computing. A like study, including time, amount, and per cent methods, and using two different intervals instead of one, might be made upon actual scores.

In the present investigation we have measured initial ability by the record made in the first trial or by the median record of the first three or the first five trials. Each method has certain advantages and certain difficulties. The second is more accurate than the first in that the record is obtained from more than one trial. This gain in accuracy may be balanced by a loss in definiteness due to the introduction of another factor—improvement in the tests. When we take into account not only the first but the second and third or perhaps the fourth and fifth trials—we are measuring not only the first attack at the problem—immediate adaptability—but the progress made during three or five trials. By employing both methods we make possible a comparison of the two and a correction of one by the other.

Final ability has also been measured in two ways by using the median of the last three trials or the best record made. The former measure is more accurate than a ranking by the last trial alone. The effect of especially high or low records at the end—due perhaps to the excitement of finishing the series or the variability occurring there are partly eliminated by this method. Learning, which would make the score of the first trial differ greatly from the median of the first three or five trials, would be of no practical importance at the end of the period of work when change is negligible. The interval between the median of the first and of the last three trials was the first basis upon which improvement was scored.

As a corresponding point to initial ability as measured by first trial, final ability as measured by the very best record wherever it occurs was chosen. The main objections to the choosing of such a point are—(1) the large part that chance may play when measurement is based on only one trial, (2) the theoretical objection that since this point is reached by different individuals at different parts of the practice curve that the amount of practice is in no sense equal—and that discussion of different amounts of improvement upon the same amount of practice is irrelevent. To the objection of possible inaccuracy we have no adequate answer. Factors such as fatigue, lack of interest, distraction during the test are mentioned by writers on practice as possible causes of inaccuracy. These would tend to make individual ability seem less than its real amount. The only chance errors which would operate to raise records beyond the ability of the individual would be errors due to apparatus or scoring methods, or "accidents in the nervous system" which might make a subject show a great proficiency, never again equaled. That the second sort of errors are less frequent than the first seems in the absence of evidence to the contrary, at least, possible. Best records might even approximate a representation of final ability when the errors of the experiment were eliminated. That such scores are fairly reliable seems to be shown by their relatively high inter-correlation (see Chapter 4.) At any rate, if they are less reliable (and if all other conditions are equal) this method will reduce all coefficients toward zero, making the positive correlations less positive and the negative less negative.

The second objection may be met by adding another record to the scores at the beginning and at the end, and by making

a change in the nature of the problem that we are discussing. The additional record is one that I have called "when reached." It is a measure of the time at which the best record is attained by each individual. The theoretical way of meeting this difficulty consists in changing our problem for the time from a discussion of improvement upon equal amounts of repetiton to a discussion of improvement based on equal opportunity for learning. Both kinds of equality, it should be recalled are included in the experiment. In the tapping and word-building tests there is an equal amount of time spent, in the other tests an equal amount of products produced. The practical aspects of the problem seem to remain almost unchanged. We give groups of subjects an equal amount of practice. The problem is not—as in the former case—who at the end of practice has attained the highest level of efficiency? but rather—who at any time in the course of the study has shown herself able to reach, by the aid of her original nature and the gain from as much practice as she may have had—a high score? A person who reaches her best score early in the practice period and whose improvement is measured on this basis, has had a very small amount of practice in comparison with her fellows at the time the measurement is made. She has, in the end, however, equal opportunity for making progress. She is not penalized by measuring her improvement on the basis of a few trials. Her record is on the contrary better than it would otherwise have been. Her lack of ability to maintain her high standing is not taken into account. Whether or not this lack of ability is a significant thing or merely an accident due to the inaccuracy of the experiment—is a point to be considered after the different measures of gain have been compared. Our two problems, based on different intervals of measurement are: "Which individuals, with equal amounts of practice, will be, at the end, at the highest level efficiency?" and "Which individuals, being given equal opportunity to dispay their ability, will best demonstrate the possibility of eventual high achievements?"

The two intervals—the difference between the median of the first three and last three trials—or the difference between the first and the best trials were used then as our basis for calculating improvement. Four methods of measuring gain were employed for both the first and the second periods. The first interval is measured by—(1) the difference between the gross

time required, (2) the difference between the gross amount of work performed in a given time, (3) the per cent which the gain measured by either of these methods was of the initial abiity, (4) or by the method of changes in ranks. An example will make these scores clear.

Let us suppose that "A" in the adding test has an initial score of 103 seconds required to add the first four columns and a final score of 79 seconds for time required to add an equally difficult four columns. The gross difference in time (first method) is 24. The gross difference in amount is obtained by transmuting the scores 103 and 79 into terms of amount done per unit of time. The difference between the reciprocals of the two numbers gives us .003. The ratio, gain divided by first trial, is different by the time and by the amount method —in one case .233 in the other .309. This difference by our scheme of correlating does not concern us—for though the actual percentage amounts may be different, their relative order by the time and amount method remain unchanged. In using the method of ranks we need to arrange all the subjects in an order of merit for initial and final ability and to use as a measure of gain the difference between the first and last positions of each individual. The person whose gain was the largest has made the largest step up on the scale—the person whose gain was the least has gone down the largest amount on the scale. All steps it should be noted, are assumed to be equal. A gain from second to first place is considered equivalent to a change from twenty-third to twenty-second positions. The value of this assumption can be tested later. In our illustrative case A—who was 20th at first became 16th in the final ranking. Her gain then was 4.

Using the second interval—difference between first and best trials—as a basis of measurement we obtain four figures —difference in time required, difference in per cent of gain, difference in relative position and difference in time at which the best record was made. The first three measures are similar to those already employed. The measure "gain in amount" was omitted, because the labor of turning time into amount seemed to over-balance the possible advantage of another measure of improvement. The possible trend of correlations by this method may be inferred perhaps by applying the results obtained when the difference between the median of the first and second trial was used, to the actual time records se-

cured from the second interval. The measure, "time at which best record is made," was obtained in the following way: The subject who made her best record earliest in the practice was considered to have made the most significant gain. In our illustration, "A" achieved her best record at the twentieth trial—her ranking is higher by this method than "B" whose best record came at the twenty-first trial.

The following tabular arrangement will give a better understanding of the measures used.

1. Basis of measurement—progress made between median of first three and last three trials.

A. Improvement measured by difference in gross time required.

B. Improvement measured by difference in gross amount done per unit of time.

C. Improvement measured by ratio, gain measured by gross methods divided by median of first three trials.

D. Improvement measured by gain in relative position.

II. Basis of measurement—progress made between initial trial and best trial.

A. Improvement measured by difference in gross time required.

B. Improvement measured by ratio, gain measured by gross method divided by first trial.

C. Improvement measured by gain in relative position.

D. Score used—time of reaching best record.

The two problems—measurement of gain and relation of improvement to initial ability may best be attacked (1) by considering the general trend of relationship, the result obtained when all the different measures of gain were correlated with the measures of initial ability, (2) by examining the agreement between the measures of gain—both by (a) comparing the correlations with initial ability obtained by using different measures of gain and (b) by comparing the tests themselves— using as a basis for this comparison, the observed relationship between original ability and improvability. Table 26 presents the data on which the conclusions are based. It shows the correlation of scores signifying improvement, with each other and with initial proficiency. In this table are included not only the results of the present investigation but a computation based upon a study made by Professor Holling-

worth in 1914.* In this experiment thirteen individuals practised seven tests, repeating each test two hundred and five times—until a definite improvement level was reached. All

TABLE NO. 26

Correlation between Initial Score and Various Measures of Gain. Investigation A—Gain Measured from Medians.

Color-naming.

	Median 1st—3	Time	Amount	%	Rank
Median 1st—3	x	-31	-02	-24	-39
Time	-31	x	93	98	87
Amount	-02	93	x	97	96
%	-24	98	97	x	96
Rank	-39	87	96	96	x

Adding.

	Median 1st—3	Time	Amount	%	Rank
Median 1st—3	x	-11	50	16	-12
Time	-11	x	78	93	91
Amount	50	78	x	91	74
%	16	93	91	x	88
Rank	-12	91	74	88	x

Multiplying

	Median 1st—3	Time	Amount	%	Rank
Median 1st—3	x	-84	30	-17	50
Time	-84	x	13	60	71
Amount	30	13	x	85	88
%	-17	60	85	x	66
Rank	50	71	88	66	x

Tapping.

	Median 1st—3	Time	Amount	%	Rank
Median 1st—3	x	-15	-09	-15	-38
Time	-15	x	99	99	89
Amount	-09	99	x	99	87
%	-15	99	99	x	88
Rank	-38	89	87	88	x

Word-Building.

	Median 1st—3	Time	Amount	%	Rank
Median 1st—3	x	-26	-05	-11	-24
Time	-26	x	88	97	92
Amount	-05	88	x	96	92
%	-11	97	96	x	90
Rank	-24	92	92	90	x

Investigation A—Gain Measured from 1st Trial to Best Trial.

Color-Naming.

	1st Trial	Time	%	Rank	When
1st Trial	x	-57	-40	-39	-11
Time	-57	x	94	80	-12
%	-40	94	x	87	-20
Rank	-39	80	87	x	-36
When	-11	-12	-20	-36	x

*For further description of the tests see Hollingworth's "Influence of Caffein upon Efficiency." Archives of Psychology, No. 22.

Tapping.

	1st Trial	Time	%	Rank	When
1st Trial	x	06	-18	-44	50
Amount	06	x	99	85	-21
%	-18	99	x	89	-27
Rank	-44	85	89	x	-41
When	50	-21	-27	-41	x

Adding.

	1st Trial	Time	%	Rank	When
1st Trial	x	-85	-68	-52	-07
Amount	-85	x	96	83	02
%	-68	96	x	91	-01
Rank	-52	83	91	x	-42
When	-07	02	-01	-42	x

Multiplying

	1st Trial	Time	%	Rank	When
Ist Trial	x	-94	-66	-48	-03
Amount	-94	x	78	61	20
%	-66	78	x	90	15
Rank	-48	61	90	x	18
When	-03	20	15	18	x

Investigation B—Gain Measured by Difference Between Median of 1st and Last 5.

Color-Naming.

	Median 1st—5	Time	Amount	%	Rank
Median 1st—5	x	-73	-29	-57	-36
Time	-73	x	81	97	83
Amount	-29	81	x	87	83
%	-57	97	87	x	90
Rank	-36	83	83	90	x

Opposites.

	Median 1st—5	Time	Amount	%	Rank
Median 1st—5	x	-98	-35	-83	-63
Time	-98	x	45	92	71
Amount	-35	45	x	66	90
%	-83	92	66	x	76
Rank	-63	71	90	76	x

Calculations

	Median 1st—5	Time	Amount	%	Rank
Median 1st—5	x	-92	-50	-69	-65
Time	-92	x	72	88	83
Amount	-50	72	x	95	94
%	-69	88	95	x	98
Rank	-65	83	94	98	x

Tapping

	Median 1st	Time	Amount	%	Rank
Median 1st	x	-03	10	-03	-50
Time	-03	x	98	99	82
Amount	10	98	x	98	86
%	-03	99	98	x	81
Rank	-50	82	86	81	x

Discrimination

	Median 1st	Time	Amount	%	Rank
Median 1st	x	-53	-53	-53	-36
Time	-53	x	1.00	1.00	78
Amount	-53	1.00	x	1.00	78
%	-53	1.00	1.00	x	78
Rank	-36	78	.78	.78	x

Crossing Test (Cancellation)

	Median 1st	Time	Amount	%	Rank
Median 1st	x	–02	27	20	–43
Time	–02	x	90	64	85
Amount	27	90	x	99	75
%	20	64	99	x	74
Rank	–43	85	75	74	x

Three Hole Test

	Median 1st	Time	Amount	%	Rank
Median 1st	x	–89	–08	–66	–32
Time	–89	x	51	89	58
Amount	–08	51	x	78	87
%	–66	89	78	x	82
Rank	–32	58	87	82	x

measurements were made in terms of time required to perform a given task. The color-naming, tapping, and calculation tests of this study are comparable to the color-naming, tapping and adding tests of the present investigation. The records here have been treated by the first method, the interval being that between the median of the first and of the last five trials. This study will be referred to as Investigation B. Table 27 is an extract from the first table, presenting some of the same facts in a different order. Correlations between measures of gain have been omitted and the relation of improvement to initial ability only is shown. An inspection of this table will give an idea of the general trend of relationship.

TABLE NO. 27

Correlation Improvement and First Records.

I. Investigation A—Measured by Median.					
Test	Time	Amount	%	Rank	When
C. N.	–31	–02	–24	–39	
Tap.	–15	–09	–15	–38	
Add.	–11	50	16	–12	
Mul.	–84	30	–17	50	
W. B.	–26	–05	–11	–24	
II. Investigation A—Measured by interval of 1st and best trials.					
C. N.	–57		–40	–39	–11
Tap.		06	–18	–44	50
Add.	–85		–68	–52	–07
Mul.	–94		–66	–48	–03
III. Investigation B—Measured by Medians.					
C. N.	–73	–29	–57	–36	
Opp.	–98	–35	–83	–63	
Calcu.	–92	–50	–69	–65	
Tap.	–03	10	–03	–50	
Discr.	–53	–53	–53	–36	
Cross-Test	–02	27	20	–43	
Three Hole	–89	–08	–66	–32	

A definite tendency of the correlations is at once apparent. The correlation of gain, measured in any of the five ways, with initial ability is negative in fifty-six cases out of the sixty-four. It is above minus 30 in one-half of the cases, above minus 10 in all but seven of the negative cases.

Exceptions to this tendency toward negative correlation are for the most part small. The positive correlations are adding (by amount) 50, (by %) 16, multiplying (amount) 30, (rank) 50, tapping (second method) 06, (Investigation B) 10, crossing test (amount) 27, (by per cent) 20. Color-naming measured by all methods in both experiments, tapping by both methods in Investigation A, word-building, multiplying by the second method, opposites, calculation, discrimination, and the three hole test show negative correlations throughout.

The preponderance of negative over positive correlations in number and amount suggests that, if our measures are accurate ones, greater relative ability at the start, in some of these tests, need not go with greater ability to improve.

A comparison of measures of gain is instructive. Table 28 gives the average correlation of each measure of improvement with the two or three other scores used. The measure called "when reached" has because it is not comparable with the other scores, been omitted here. So improvement in color-naming, measured by the time method has an average correlation of .93 with improvement as measured by the amount, per cent, and rank method. Next to each entry in the table is a figure in parentheses (one, two, three, or four) which represents the rank of the average correlation as compared with other average correlations in the same test by different methods. The (4) next to 93 (time) in color-naming means that in this test (medians being used as measures) the average correlation of time with the other three measures was found to be less high than the average correlation of any one of the other measures with all three. The (1) after per cent shows that per cent had the highest average correlation. Amount and ranks came next.

In this table, as we would expect, all correlations are positive. They are above fifty per cent except in two cases, above sixty except in five cases. Rankings by different measures of gain seem, then, to resemble each other.

The measures have a certain individuality. In all the experiments together per cent has the highest correlation with

other measures of gain—then come in order, ranks, amount, and time. The ranking by per cents is the same—it should be noted—whether the percent is obtained from a measure of amount of time taken or amount of work performed. If one wishes an order of gain which will be the same whichever way it is obtained and which will indicate most closely the results given by other measures, per cent seems to be the one to use. On the other hand, as has been pointed out many times, this figure has the disadvantage of being further removed from the original results than are figures of actual amount of work done or time required.

TABLE NO. 28

Average Correlation of Measures of Gain.

Test	Time	Amount	%	Rank
I. Investigation A—Median used as measure.				
Color-naming	93 (4)	95 (2)	97 (1)	93 (3)
Adding	87 (2)	81 (1)	91 (1)	84 (3)
Tapping	95 (3)	95 (2)	96 (1)	88 (4)
Multiplying	48 (4)	62 (3)	70 (2)	75 (1)
Word-building	92 (2)	92 (3)	94 (1)	91 (4)
II. Investigation A—1st and Best used as measure.				
Color-naming	87 (2)		91 (1)	84 (3)
Adding	89 (2)		94 (1)	87 (3)
Tapping		92 (2)	94 (1)	87 (3)
Multiplying	69 (3)		84 (1)	76 (2)
III. Investigation B—Median used as measure.				
Color-naming	87 (3)	84 (4)	91 (2)	95 (1)
Opposites	70 (3)	67 (4)	78 (2)	79 (1)
Calculations	81 (4)	87 (3)	93 (1)	92 (2)
Tapping	93 (1)	92 (2)	91 (3)	83 (4)
Discrimination	93 (2)	93 (2)	93 (2)	78 (4)
Crossing	80 (2)	88 (1)	79 (3)	78 (4)
Three Hole	66 (4)	72 (3)	83 (1)	75 (2)

The difference between measures of gain can be demonstrated further by a consideration of Table 27 which gives the correlation of different measures of gain with initial ability. This correlation is, except in three cases where they are equal, less strongly negative in the case of per cent than of time. Similarly correlations by amount are more positive than correlations by per cent. The reason for this is inherent in the measures themselves.

A hypothetical case will demonstrate this. Let us suppose that A and B took 25 and 50 seconds respectively to perform a certain task. A, by the time method is better at the beginning than B. Let each one make a gain of five seconds—by the time method an equal gain. If we change these gains into per cents, the equality is lost. A, who was best at first has gained

20%—B who was less skillful has gained only 10%. The initially better person by the per cent method, is given the advantage. So the tendency toward negative correlation, between initial ability and improvement, so marked in the case of the time measure, is reduced by a change to the per cent method, which, when gains equal by the time method are compared allows the person with the better initial record to make the larger gain. The small size of the negative correlation measured by the amount method can be demonstrated in the same way. Turning the original 25 and 50 into terms of amount accomplished—we get 04 and 02. A and B by the original hypothesis each required five seconds less at the end of the practice period. Their times were then 20 and 45 seconds. These turned into amounts give 05 and 022. The one who was better at the beginning is—as in the case of per cent—given the advantage.

But this advantage is greater than the advantage given by the per cent method. In the example the gain of the better person when measured by the per cent method is twice the gain of the poorer, when measured by the amount method, however, this relation is five to two. Amount then will have the least high negative correlation as shown by the fact that the better person, by this method, has the greatest advantage over the poorer. Gains made by persons good at the beginning are raised by a change to the amount method, the negative relation between initial ability and improvability consequently reduced.

Objections are easily raised against the measure which has been called "Rank." Its obvious fault is that it allows for no improvement for the person who is best at first and no loss for the person who is initially worst. Furthermore, rise in position from sixth to third place, for example, is considered equal to a change in rank from 22nd to 19th. Another difficulty was noted when the results were correlated. Almost all gains in rank were small and there were frequently three or four or even five which were equal in amount. Correlations by the rank method are not as accurate when a number of the quantities in one of the orders are equal as when the differentiation of positions is obvious. The rank method when tried out, however, showed a surprisingly high correlation with other measures and a low negative correlation, as did other methods, with initial ability. The evidence from the experiment seems

to show that if the other three scores are valid, that this also gives fairly truthful result. In the present experiment it happened infrequently that the person who was first, remained first, or the person who was last, remained last, so the question of undistributed extremes did not affect the results materially. The assumption that gains at the ends of the scale are equal is one which underlies the other measures as well and so will not affect the correlations between the methods of scoring improvement. It may be argued in favor of this way of measuring, that in daily life gain is more frequently measured in terms of relative than of absolute achievement. The rank method makes improvement contingent on the achievement of others as well as upon the subject's own absolute gain.

The measure called "when reached" is not a measure of improvement in the sense that the others are records of progress. It correlates negatively with other measures of gain, except in the case of adding, a correlation of +02, and in the case of multiplying where there is a small positive correlation throughout. With one exception (tapping) the correlation between initial ability and "when" is low and negative. There seems to be no relation between reaching one's maximum early and being proficient at the start, and a tendency toward negative relation between achieving one's best record early and making a large amount of gain. This latter is a correlation which seems on the face of it, to be a probable one. but, as shown by the small size of the coefficients, not inevitable.

In observing the results gained by the different methods one notes that in three cases out of four the correlation between initial ability and improvement as measured by difference between best scores and initial scores is more strongly negative than the correlation between initial ability and improvement measured by the difference between the medians of the first and last trials. The exception noted is in the case of tapping in whch the two coefficients are approximately equal. Best records and initial records would seem from their derivation to be less reliable than median records. But chance errors would operate to reduce coefficients toward zero rather than toward minus 100. A constant error might be operative here. If it were true that best records always tend to raise a score above its true amount (or first records to lower it unduly), then since there is greater opportunity in the case of the initially poor than of the initially good individuals, such

an error would make the change made by those poorest at the start seem greater than the change made by the initially better people. This error, on the other hand, is only the error to which all unweighted measurements of improvement are subject. We are not discussing the question of opportunity, difficulty, probability, but of actual observable amounts of gain. It is improbable also that best records were frequently much accelerated. Where such increase was due to an "accident in the nervous system" it cannot be called an error, for we are inquiring here not into average performance but ultimate capacity, however caused. The lowering of first records is likewise, if present, an actual fact which should be included in the measurement. The fact that best records have a high intercorrelation and a low negative correlation with improvement, (as do other measures) would seem to argue for their reliability. A possible explanation of the higher negative correlation of best scores might be the fact that the first trial —best trial interval allows for the greatest amount of improvement whereas the median method has, as seen above, irrelevantly admitted opportunity for learning.

Besides the relation between first records and improvement and the relation between the intervals and scores which measure gain, the differences in the tests themselves may be studied—when considered from the point of view of the relation between gain and original ability. We can arrange the tests in a hierarchy using as the basis for our ranking the amount of negative correlation between initial ability and improvability. We may use the "time" method as the measure of gain—as this was the one in which all but three of the original measurements were made. Then we have Table No. 29. In Table No. 30 the arrangement is made by an averaging of the correlations by time, per cent, and amount (where amount was calculated).

That something is actually measured by this hierarchy might be argued from the fact that the same tests by different measures or in different experiments so frequently fall relatively near together in these tables. We have color-naming in Table 29 consistently high –73, –57, –31. Tapping in the same table is always low –15, –03 and 06. Near the tapping test is the crossing tests (–02) which seems to be somewhat similar.

Multiplying has a high negative correlation when measured by either method. Near it are opposites, calculation, and add-

TABLE NO. 29

Negative Correlation of First Records with Improvement as Measured by Time.

Investigation		
B	Opposites	-98
A 2nd Method	Multiplying	-94
B	Calculation	-92
B	Three Hole	-89
A 2nd Method	Adding	-85
A 1st Method	Multiplying	-84
B	Color-naming	-73
A 2nd Method	Color-naming	-57
B	Discrimination	-53
A 1st Method	Color-naming	-31
A 1st Method	Word-building	-28
A 1st Method	Tapping	-15
A 1st Method	Adding	-11
B	Tapping	-03
B	Crossing	-02
A 2nd Method	Tapping	06

TABLE NO. 30

Hierarchy of Negative Correlation of First Records—Measures Averaged

Investigation		Average
A Method 2	Multiplying	-80
A Method 2	Adding	-77
B	Opposites	-72
B	Calculations	-70
B	Three Hole	-54
B	Discrimination	-53
B	Color-naming	-53
A Method 2	Color-naming	-49
A Method 1	Multiplying	-24
A Method 1	Color-naming	-19
A Method 1	Word-building	-14
A Method 1	Tapping	-13
A Method 2	Tapping	-06
B	Tapping	01
B	Crossing Test	15
A Method 1	Adding	18

ing (by one method). Adding and calculation differ of course only in the particular form of the operation involved. The similarity between the opposites and the calculation tests has frequently been commented upon. Table No. 30 shows (with some exceptions) a like result. Tapping is consistently low—and ranks with crossing—color-naming is consistently mediocre, multiplying (by one method), opposites and calculation are all high. The discrimination reaction and the three hole test rank with these.

A general tendency is brought out by the two hierarchies. Tests of the higher, "more intellectual" functions seem to gravitate toward the top of the list—those of the lower

"motor" functions are more frequently found below the middle. Opposites, calculation, and multiplying are definitely at the top. Tapping and crossing are always at the bottom, and color-naming is always mediocre. Were the order strictly that of intellectual to less intellectual functions probably the position of word-building would be raised and that of discrimination lowered.

The tendency is sufficiently well marked for us to observe that in tests employing higher mental functions, initial proficiency is less likely to imply subsequent improvement than is ability at the start in lower forms of achievement likely to indicate probable gain to be made in those particular tests.

It is difficult to explain this hierarchy. Closeness of approach to a limit, or relative unreliability of the more complex performances are possible though not probable explanations. Tests at the top of the hierarchy show no consistent tendency to be those in which the end of improvement is most closely approximated. Opposites and calculations, it is true, showed a level of improvement but tapping (Investigation B) did also. Color-naming which showed a definite level stands, not as we might expect, at the top but in the middle. Multiplying —the test in which there was the greatest possibility of gain of all—is not at the bottom but near the top of the list.

Nor does the hierarchy follow the order of reliability. The reliability of the tests in Investigation B is not determined but because of the large amount of practice given it is safe to assume that individual variability has been greatly reduced. Calculations of Investigation B should then be more reliable than adding of Investigation A. It is equally or more strongly negative. Similarly color-naming in Investigation B, with a larger practice period is more negative than color-naming in investigation A. The color-naming and adding tests of the present study contrasted with the more variable multiplying, tapping and word-building tests. Multiplying which is most unreliable is at the top of the list, tapping also unreliable is at the bottom. If multiplying were excluded one might venture the suggestion that the least reliable tests had their coefficients reduced toward zero, whereas in those that are less variable, in which practice is continued for the longest time—the correlations are more strongly negative. All arguments from such a hierarchy are dangerous, not only because of the number of factors involved, but because of the chance errors

which may enter in, the tendency to argue from striking cases, etc. It is sufficient to observe that this hierarchy seems to give results opposite to those obtained by other investigators—showing that improvement in the more complex functions has a larger negative relation to original skill than has improvement in the simpler functions.*

Since the matter of unreliability may still be urged as an objection it might be well to determine improvement by using two of the least variable measures even though this determination reduces enormously opportunity for gain. The practice series (of Investigation A) for this purpose, was divided into halves. Eleven trials were included in each half. Improvement is measured from the median of the first to the median of the second half. Table 31 shows the correlation between the first and second halves, and also the correlation between initial ability, as measured by the median of the first eleven trials, and improvement, as measured by the difference between the median of the first and second halves.

TABLE NO. 31

	Cor. Initial Ability and Improvement	Cor. between two Halves
Color-Naming	–21	.96
Adding	–60	.94
Tapping	34	94
Mul.	–54	87
Word-Building	–40	88

The correlations here are still, with the exception of tapping, negative. They are less in the case of multiplying and color-naming, very much greater in the case of adding. The reduction of the coefficients toward zero is explicable as a result of reducing the amount of practice. A rise in correlation toward minus one hundred might be due to increased reliability of results.

The very high correlation between the two halves of the test is important. In spite of the fact that the poorest individuals are gaining the most—whether gain is measured in terms of actual amount or time gained—they do not, during the interval between the median of the first and of the last

*If the hierarchy were in terms of amount done the order would be different. But as this is the method furthest removed from the original measures such a hierarchy would seem to be of lesser value. What the results would be had the original scores been in terms of amount done and the practice equated in terms of time spent we do not know.

half of the experiment, catch up with those who were initially most proficient. Changes in rank are more noticeable (see Chapter 2) when the median of three trials or one trial only is considered.

In order to make this study comparable to others on the same subject, Table 32 is added, which gives the average gain for the six who were initially most proficient, the five who were least skillful, and the two intervening sets of six. Improvement is measured by the time method over the interval, median of the first to median of the last three trials.

TABLE NO. 32

Indiv.	C. N.	Tap.	Add.	Mul.	W. B.
1-6	11.63	31.1	22.2	101.8	4.65
7-12	12.16	64.5	21.3	403.8	7.0
13-18	14.63	33.8	18.5	623	8.1
19-23	36.69	103.6	27.8	738.6	6.5

Averages point to the same result as did the correlations. The initially better gain the least. Those variations which are observed in this table might easily be chance results due to the small number of cases in each group. If the first two and the last two groups are averaged there is a perfect correspondence between original proficiency and little gain.

The correlations and averages, then, seem to show a general tendency toward negative relationship between initial ability and improvement through practice. This is true whether we measure the gain over long or short intervals of practice, whether in general, we use the time, amount, per cent, or rank method, whether the function studied is the complex multiplying or the simpler color-naming. In general it seems true that the longest intervals and in some cases the apparently more reliable measures give the highest negative correlation. The time method is always most strongly negative, the per cent next, and amount last. The more complex functions show this tendency more markedly than do the simpler traits.

The limits of diagnostic value of the initial or of the first two or three trials of a test are immediately obvious. Such tests are indicative (see Chapter 3) of final status. They are of little value in determining the actual amount or per cent of improvement an individual will make during a given amount of practice. An initial high score should make one predict relatively little rather than relatively great future gain.

CHAPTER FIVE

THE RELATION OF IMPROVEMENT TO FINAL ABILITY.

Closely related to the problem of initial achievement versus probability of improvement is the question of the correlation of final ability and amount of gain. The problem here is—what is the cause of the final high ranking of some persons, and the low position of others? Have those who are eventually best achieved their score through improvement in the practice series and have those who are poor shown a notable lack of gain—or must some other factor be suggested to explain final position?

The answer to this question can be found most easily by correlating the final order with the order of gains. As in the previous study, two intervals were used upon which to reckon gain. Two scores for each of these intervals, the original gross improvement and the percentage of improvement were correlated with final ability. "When reached" the measure of the time at which the maximum record was made, was also correlated with final achievement. Table No. 33 presents the results of this correlation.

TABLE NO. 33

Final Position and Amount of Improvement.

	By Time	%	When
Color Naming			
Final Position & Gain	18	27	
Best Score & Gain	–03	23	–41
Tapping			
Final Position & Gain	64	61	06
Best Score & Gain	70	62	
Multiplying			
Final Position & Gain	–07	.66	
Best Score & Gain	–34	23	11
Adding			
Final Position & Gain	27	51	
Best Score & Gain	08	34	–36
Word-Building			
Final Pos. & Gain	48	39	
Experiment B.			
Color-naming	–03	18	
Opposites	–02	26	
Calculation	21	62	
Tapping	87	91	
Discrimination	18	18	
Crossing Test	72	86	
Three Hole	–77	–26	

In this table there is a preponderance of positive over negative coefficients. Of the 32 figures given there, excluding "When reached," only seven are negative. Except in the case of the three hole test all correlations, when gain is measured by the per cent method, are positive. Except in four cases, the determination by per cent has a higher correlation than the determination by original measures. In one of these cases both have the same correlation, in three the original record was made in terms of amount. It can be demonstrated by the means applied in Chapter 4 for a similar purpose, that the best person at the end is given a greater advantage in reckoning improvement by the amount than by the per cent method, and by the per cent than by the time method. So we should expect correlations by amount to be higher than by per cent which are similarly higher than by time.

This table, then, shows a general tendency toward a positive relation. Ability to improve by practice seems to be a factor in the gaining of a final high position. This ability to improve is best defined, however, by considering it as the capacity to make a gain which is a large per cent of one's original score, or to perform a greater amount of work in a given time rather than as ability to perform a given amount of work very much more quickly. The relative size of the correlation by per cent emphasizes the importance of considering gain—when we speak of its correlation with final position—as an improvement large in relation to one's original score. The finally best person in multiplying does not make a gain of a greater number of seconds than the worst person. Because of the wide range of scores this would be in some cases impossible. But he does gain a large per cent of his original score.

It is important to note that the best score in three cases out of four correlates less closely with improvement than the last score. The exception noted is tapping, which was originally measured by the amount method. The two coefficients are in this case approximately equal. The three other cases seem to show a tendency for the person who is best at the end to owe her position largely to practice gains—while the person who attains a high ranking somewhere in the practice series seems to derive this ability not so surely from improvement made between first and best score. This is, possibly what one would expect. He who is best at the end has had a longer time in which to make large practice gains. The person who is most

efficient when efficiency is measured by a high score made anywhere in the practice period may have made his best record very early in the game and may, with less opportunity for improvement by practice, be depending more upon his native ability. Such a finding is similar to that observed in Chapter 4, where the correlation of gain, measured by initial score—best record interval, also gave a higher negative correlation with initial records. The negative correlation of reaching one's best score early in the series, found in color-naming and adding contrasts with the lack of relation observed in tapping and multiplying. The finally best person in color-naming and adding seems to continue improving for a longer time than other less skillful individuals. In tapping and multiplying there seems to be no correlation between final achievement and time of reaching one's improvement level.

In order to check the results obtained by these two methods, the correlation was obtained between final position, as measured by a very reliable score—the median of the last eleven trials, with gain as measured by the difference between the median of the first and of the last eleven trials, in Experiment "A". Table 31 (see Chapter 4) is repeated here—with the additional middle column.

TABLE NO. 34

	Gain & Median 1st Eleven Trials	Gain & Median Last Eleven Trials	First & Last Eleven Trials
C. N.	-21	-03	96
Tap.	34	54	94
Add.	-60	-25	94
Mul.	-54	-16	87
W. B.	-40	04	88

The correlation between gain and final position is less strongly negative than the correlation between improvement and initial rank. This result is partly due to the mathematics of measuring gain. If we call the original score, a, the gain, g, then final score will equal a+g. Because of the common element g, a+g will be more likely to have a positive correlation with g, than will some other figure.

Improvement measured over this interval shows a tendency, with the exception of tapping, toward a negative or at least toward zero correlation—whereas gain measured over the two longer intervals seemed to be related positively to final score. From such a comparison we must, of course, omit the

scores in terms of per cent and consider only those gains based directly on original records. Though the correlation of gain by the first two methods is not always indisputably positive, it is usually nearer the positive end of the scale than is the correlation by the third method. The order of positiveness is,—first, correlation with gain measured over the interval, median of the first three to median of the last three records, then correlation with gain measured over interval initial score to very best score, and finally gain measured over the interval median of the first eleven to median of the last eleven trials. These intervals differ in reliability and in length. We may say that over the longer, less reliable intervals the relation as measured is frequently positive, over the shorter intervals with more precisely determined score, the relation is more frequently a negative or zero one.

The exception in the case of tapping is notable. In both investigations "A" and "B" final position is seen to be correlated positively with practice gain. One cause for this correlation may be the fact that tapping was measured by the amount method. Or there may be some reason inherent in the test itself. A hierachy of positive correlations (based on original scores measured over the first and second interval) may show something about the difference between the tests.

TABLE NO. 35

Hierachy—Gain correlated with final position—Time used as Measure.

Tapping (B)	.87
Crossing	.72
Tapping (A) (2nd. method)	.70
Tapping (A) (1st Method)	.64
Word-building (1st Method)	.48
Adding (A) (1st Method)	.27
Calculation	.21
Color-naming (A) (1st Method)	.18
Discrimination	.18
Adding (A) (2nd Method)	.08
Opposites	−.02
Color-naming (A) (2nd Method)	−.03
Color-naming (B)	−.03
Multiplying (A) (1st Method)	−.07
Multiplying (A) 2nd Method)	−34
Three Hole	−77

The significance of the hierachy is not very plain. It shows a certain tendency toward an opposite ranking to that of the arrangement in Chapter 5. Tapping—at the bottom before—is at the top here. Opposites is near the bottom instead of the

top. Color-naming which was in the middle, nearer the top in the former arrangement is now nearer the bottom. Discrimination has changed its position markedly. Word-building, multiplying and calculation remain about the same. With the exception of the three hole test and of word-building we might say that the correlation, on the whole, grows less as we approach the more complex, more strictly intellectual functions. These might also be called the less reliable functions.

It is difficult to draw a conclusion from such diverse results. One may emphasize a few observed facts. These are (1) a definite positive correlation between final score and gain when per cent or amount measures are used, (2) a tendency toward postive relationship even when time measures are used if the intervals are sufficiently long, (3) a reduction of this relation toward zero or toward negative correlation when very accurate measures are used at either end of a short practice period, (4) a tendency for the simpler functions to show a stronger positive correlation. One might expect, then, that long continued practice at some relatively simple task, would give, especially if gain were measured in terms of per cent of original score or difference in amount of work performed, a high positive correlation between final position and improvement. More difficult performances, especially if repeated for a short time only and measured in terms of time required might show a small positive relation, no relation or a negative correlation between gain and rank at the end of the series of repetitions.

Of greatest importance in all these discussions are the figures in Tables 3, 4 and 5, (Chapter 2) and Table No. 34 (Chapter 5). Though the relation of final position and gain may vary with accuracy, length of practice, kind of functions studied, the correlation between initial and final position is always a positive one.

CHAPTER SIX

THE RELATION OF IMPROVEMENT IN ONE TEST TO IMPROVEMENT IN ANOTHER AND OF INITIAL AND FINAL GAIN.

The correlation between scores in the different tests and between initial and final position has been considered. Problems of relation may be studied not merely by means of a comparison of individual records but through the correlation of individual improvement. We may ask, does gain in one test imply improvement in one or all of the other tests? Is gain at the beginning of the practice series prognostic of gain at the end?

In order to make the measure of improvement as reliable as possible the amount of practice has been reduced, in obtaining an answer to our first problem, and that measure of gain employed which is the difference between the medians of the first and of the last eleven trials. No correction for attenuation has been made. Correlations of improvement in every test with improvement in every other test give the followng table:

TABLE NO. 36

Correlation of Improvement in Every Test with Improvement in Every Other Test.

	C. N.	Tap.	Add.	Mul.	W. B.
C. N.	x	-.08	-07	-04	18
Tap.	-08	x	12	-21	-18
Add.	-07	12	x	-09	15
Mul.	-04	-21	-09	x	-04
W. B.	18	-18	15	-04	x
Average	-003	-09	03	-10	03

The relation according to this measure of gain, in these particular tests, seems to be purely a chance one. Even such similar tests as adding and multiplying have a correlation which is very nearly zero. There is no evidence here for any general capacity for improvement.

The correlations are similar to those obtained by Chapman and Wimms. The average inter-correlations between Chapman's tests measured by two improvement scores were, for color-naming, adding and multiplying:

TABLE NO. 37

	C. N	Add.	Mul.
C. N.	x	–11	.00
Add.	–11	x	.05
Mul.	.00	–05	x

Wimms** gets a correlation of .007 between improvability in addition and improvability in multiplication.

As Chapman points out* our practice periods are probably too short. Improvement measured from infancy up would probably give a positive correlaton. But over such periods of time as we have been able to measure gain and with such scores as we can obtain, there is no evidence whatsoever of a positive relation between gain in one test and gain in another. This fact contrasts with the discovery that even first trials of tests frequently show a positive correlation. What a person can do in one test is a better indication of what he can do in another, than is his ability to improve in one capacity a sign of ability to make progress in another.

The problem of the correlation between early and late gain is an important one. It might be that gain during the first part of an intensive practice experiment would be evidence of a capacity for improvement along that particular line, so that the individual able to make a larger improvement than his fellows at the beginning of practice would also make larger relative gains at the end of the series. On the other hand it is possible that large improvement at the beginning would, because of the approximation of a physiological limit, preclude large gains at the end, and we would find a negative correlation between initial and final gain. An examination of the relation between gain measured at different parts of the practice period might throw some light upon the value of any measure of improvement as an index of an individual's ability for further improvement.

Correlations between gain in the first part of the experiment (measured by the difference between the medians of the second and fourth divisions) and gain in the last part of the series (measured by the difference between the medians of the sixth and of the eighth divisions) make up the followng table:

*Chapman—Individual Differences in Ability and Improvement, T. C. Contrib.

**Wimms—The Relative Effects of Fatigue and Practice B. J. P. vol. 2.

TABLE NO. 38

Correlation between Initial and Final Improvement.

Test	Correlation
Color-naming	.12
Tapping	−.23
Adding	.04
Multiplying	−.09
Word-building	−.15
Average	−.06

Test	Early & Late Improvement
Color	.19
Addition	.40
Mental Multiplication	.40

The correlation, again, seems to be negative or insignificant. It is noticeable that the less reliable tests give, in general, a higher negative correlation. Chapman's results for the same tests give a positive coefficient.

The color-naming test has a correlation approximately equal to the one found here.

The main difficulties with the scores which have been used are:

I. Individuals start at different points having had more or less previous practice and gained more or less from that experience. 2. Where a physiological limit is reached or approached correlations, after this point is attained by any individual are affected. 3. Our measures of gain even throughout the period of improvement may not be equitable. 4. Our actual scores are frequently unreliable. The coefficients which we have obtained are caused, we believe, by a combination of these factors. They give no evidence for the fact that ability for continued improvement may be possessed by individuals who begin by improving. Nor do they show that gain at first necessarily excludes gain at the end. The chance relationship may be a compound of the two tendencies.

One fact is evident. Improvement measures are of little value for diagnosis. They do not indicate either ability to improve in related tests or ability to improve over a different period of time in the same test. These measures are decidedly inferior in their reliablty and their amount of relationship with other scores to such measures as initial or final records. The former are measures of difference. They depend not only upon proficiency but upon a lack of ability. For an improvement score to be very large, one of the scores from which it is derived must be a poor one. In this, of course, they resemble

"fatigue" scores* and the problem is also analogous to that met with in determining whether "quick learners" are "long retainers." But actual records and average records depend purely upon proficiency. Whereas learning tests, measures of improvements over short periods of time may give no insight into general or special ability to gain, actual records in one test may be indicative of scores to be obtained in another test and, in this experiment, are diagnostic of future rank in the same test.

*For a short discussion of this see Wells' "A Neglected Measure of Fatigue." A. J. P. 1908.

CHAPTER SEVEN

SUMMARY

The results of this study summarized in what seems to be the order of their importance and definiteness are:—

1. Correlations between functions become more positive as those tests become better samples of the individual's true ultimate capacity.

This is demonstrated by the increase in intercorrelations which continues as long as there is improvement through practice. That this change is not entirely due to decreased variability of the indvidual scores is suggested by—

a. The fact that correction of the coefficients for attenuation does not do away with the increase.

b. The very high inter-correlation of those records which best represent ultimate capacity (very best records made in the entire series) in spite of their presumable "unreliability."

2. Correlations between records made in the same function at different parts of the practice series are always positive and usually high. Even beginning and end records show a positive correlation.

3. Correlations between improvement and initial records are usually negative. The longer the interval allowed for gain and the more complex the functon, the higher becomes the negative correlation.

4. Final position, especially when the per cent or amount measure of gain is used, is frequently positively related to improvement. The simpler functions show this more markedly.

5. Correlations between gains made in different functions are only such as might occur by chance.

6. No relation is found between improvement in one part of the practice period and improvement in another part of the practice period, of the same function.

A few minor results are:—

1. Correlation between alternative measures of gain over the same period in the same function are postive and high.

2. Gain measured by the "time required method" correlates more negatively with initial position than does gain

measured by the per cent method which is similarly more negative than gain measured by amount done.

3. The best of three records is, in general, as reliable a score as is the median of those records.

4. A correlation of adjacent scores in the same test confirms the results of other observers. Individual variability decreases as long as improvement in the functions continues.

The implication of these results seems to be that records made near the beginning of practice are indicate of future proficiency in the same function although initial ability may prove to be not diagnostic of but rather opposed to further improvement. Scores made by an individual in one performance are diagnostic of his relative standing in other performances, particularly if the records used are those obtained after a certain amount of improvement through practice has occurred. The reliability of actual scores is contrasted with the lack of agreement with other measures, of records of gain. Gain in one performance bears no relation to gain in another, nor is gain at the beginning of the practice series indicative of further improvement.

We find here then, another demonstration of the positive correlation existing between desirable traits. Our results may be used to show that individual differences decrease through practice, though various difficulties encountered in measuring improvement and the high correlation observed between initial and final position, may indicate that the true conclusion is the opposite of the one that the statistical procedure seems to demonstrate.

A few suggestions for further investigation of this problem might be offered. The precautions to be emphasized are those which provide for the certainty of change and those which prevent the penalizing of the initially better individual.

1. The functions studied should be sufficiently complex so that there is opportunity for improvement. The tapping and color-naming tests of the present experiment are, though they give reliable results, undesirable for the purpose because of the small amount of improvement observed. Mental multiplication would be a good choice for such a study.

2. The amount of practice given should be greater than given in former investigations. Different results are observed even in the present study when gain is measured over differ-

ent intervals. An extension of the repetition period to still greater lengths might change the conclusions.

3. Gain should preferably be measured in terms of amount of work accomplished in a given time.

4. The group should consist, in so far as possible, of unselected individuals.

MOOD IN RELATION TO PERFORMANCE

BY

ELIZABETH T. SULLIVAN, Ph. D.

Stanford Library
AUG 12 1923

ARCHIVES OF PSYCHOLOGY
EDITED BY R. S. WOODWORTH
No. 53

NEW YORK
MAY, 1922

AGENTS: G. E. STECHERT & CO. : London (2 Star Yard, Carey St., W. C.); Paris (16, rue de Condé)

ARCHIVES OF PSYCHOLOGY

SUB-STATION 84, NEW YORK CITY. The Subscription price is five dollars per volume of 500-600 pages. Volume I comprises Nos. 1-10, Volume II, Nos. 11-18, Volume III, Nos. 19-25, Volume IV, Nos. 26-32, Volume V, Nos, 33-39, Volume VI, Nos. 40-46, and Volume VII, Nos. 47-52. The available numbers are as follows:

2. On the Functions of the Cerebrum: The Frontal Lobes: SHEPHERD IVORY FRANZ. 50c.
3. Empirical Studies in the Theory of Measurements: EDWARD L. THORNDIKE. 50c.
4. Rhythm as a Distinguishing Characteristic of Prose Style: ABRAM LIPSKY. 50c.
5. The Field of Distinct Vision: W. C. RUEDIGER. 70c.
6. The Influence of Bodily Position on Mental Activities: ELMER E. JONES. 50c.
7. A Statistical Study of Literary Merit: FREDERIC LYMAN WELLS. 80c.
8. The Relation between the Magnitude of the Stimulus and the Time of Reaction: SVEN FROEBERG. 35c.
9. The Perceptual Factors in Reading: FRANCIS MARION HAMILTON. 50c.
10. Time in English Verse Rhythm: WARNER BROWN. 70c.
11. The Hearing of Primitive Peoples: FRANK G. BRUNER. $1.00.
12. Studies in Development and Learning: EDWIN A. KIRKPATRICK. $1.00
13. The Inaccuracy of Movement: H. L. HOLLINGWORTH. 80c.
14. A Quantitative Study of Rhythm: HERBERT WOODROW. 60c.
16. On certain Electrical Processes in the Human Body and their Relation to Emotional Reactions: FREDERIC LYMAN WELLS and ALEXANDER FORBES. 40c.
17. The Relative Merit of Advertisements: EDWARD K. STRONG, JR. $1.00.
18. Attention and Movement in Reaction Time: J. V. BREITWIESER. 50c. (Cloth, 75c.)
19. An Empirical Study of Certain Tests for Individual Differences: MARY THEODORA WHITLEY. $1.25. (Cloth, $1.50.)
20. Visual Acuity with Lights of Different Colors and Intensities: DAVID EDGAR RICE. 50c. (Cloth, 75c.)
21. The Curve of Forgetting: C. H. BRAN. 45c. (Cloth, 70c.)
22. The Influence of Caffein on Mental and Motor Efficiency: H. L. HOLLINGWORTH. $1.50. (Cloth, $1.75.)
23. Reaction Time to Retinal Stimulation. A. T. POFFENBERGER, JR. 70c. (Cloth, 95c.)
24. Interference and Adaptability: ARTHUR JEROME CULLER. 75c. (Cloth, $1.00.)
25. Reaction to Multiple Stimuli: JOHN WELHOFF TODD. 60c. (Cloth, 85c.)
26. A Study in Incidental Memory: GARRY C. MYERS. $1.00. (Cloth, $1.25.)
27. A Statistical Study of Eminent Women: CORA SUTTON CASTLE. 80c. (Cloth, $1.05.)
28. The Mental Capacity of the American Negro: MARION J. MAYO. 60c. (Cloth, 85c.)
29. Experimental Studies in Judgment: H. L. HOLLINGWORTH. $1.25. (Cloth, $1.50.)
30. The Psychological Researches of James McKeen Cattell: A Review by Some of His Pupils. $1.00. (Cloth, $1.25.)
31. Fatigue and Its Effects upon Control: ISAAC EMERY ASH. 60c. (Cloth, 85c.)
32. The Transfer Effects of Practice in Cancellation Tests: MELVIN ALBERT MARTIN. 60c. (Cloth, 85c.)
33. The Intellectual Status of Children Who are Public Charges: J. L. STENQUIST, E. L. THORNDIKE, and M. R. TRABUE. 50c. (Cloth, 75c.)
34. The Relation of Quickness of Learning to Retentiveness: DARWIN OLIVER LYON. 50c. (Cloth, 75c.)
35. The Overcoming of Distraction and Other Resistances: JOHN J. B. MORGAN. 75c. (Cloth, $1.00.)
36. The Psychology of the Negro—An Experimental Study: GEORGE OSCAR FERGUSON, JR. $1.25. (Cloth, $1.50.)
37. The Effect of Distraction on Reaction Time: JOHN E. EVANS: $1.00. (Cloth, $1.25.)
38. The Effect of Humidity on Nervousness and on General Efficiency: LORLE IDA STECHER. 90c. (Cloth, $1.15.)
39. The Mechanism of Controlled Association: MARK A. MAY. 75c. (Cloth, $1.00.)
40. Recitation as a Factor in Memorizing: ARTHUR I. GATES. $1.00. (Cloth, $1.25.)
41. Mental Fatigue during Continuous Exercise of a Single Function: THOMAS RUSSELL GARTH. 85c. (Cloth, $1.10.)
42. A Psychological Study of Trade-Mark Infringement: RICHARD H. PAYNTER, JR. 85c. (Cloth, $1.10.)
43. Individual Differences and Family Resemblances in Animal Behavior: HALSEY J. BAGG. 70c. (Cloth, $1.00.)
44. Experimental Studies in Recall and Recognition: EDITH MULHALL ACHILLES. 90c. (Cloth, $1.25.)
45. The Morphologic Aspect of Intelligence: SANTE NACCARATI. 70c. (Cloth, $1.10.)
46. Psychological Examinations of College Students: F. EDITH CAROTHERS. $1.25.
47. The Effects of Practice on Judgments of Absolute Pitch: EVELYN GOUGH. $1.25
48. An Experimental Study of Silent Thinking: RUTH S. CLARK. $1.40.
49. Some Empirical Tests in Vocational Selection: HERBERT W. ROGERS. 75 cents.
50. Adenoids and Diseased Tonsils: Their Effect on General Intelligence. MARGARET COBB. $1.00.
51. An Experimental Study of the Factors and Types of Voluntary Choice. ALFRED H. MARTIN. $1.50.
52. Some Well-Known Mental Tests Evaluated and Compared. DOROTHY RUTH MORGENTHAU. 80 cents.
53. Mood in Relation to Performance. ELIZABETH T. SULLIVAN. $1.00.

Other monographs to be obtained from the ARCHIVES OF PSYCHOLOGY include the following:

Measurements of Twins: EDWARD L. THORNDIKE. 50 cents.
Avenarius and the Standpoint of Pure Experience: WENDELL T BUSH. 75 cents.
The Psychology of Association: FELIX ARNOLD. 50 cents.
The Measurement of Variable Quantities: FRANZ BOAS. 50 cents.
Linguistic Lapses: FREDERIC LYMAN WELLS. $1.00.
The Diurnal Course of Efficiency: HOWARD D. MARSH. 90 cents.
The Time of Perception as a Measure of Differences in Sensations: VIVIAN ALLEN CHARLES HENMON. 60 cents.
Interests in Relation to Intelligence. LOUISE E. POULL. 80 cents.

THE JOURNAL OF PHILOSOPHY

There is no similar journal in the field of scientific philosophy. It is identified with no philosophical tradition, and stands preëminently for the correlation of philosophy with the problems and experience of the present.

The contents of recent numbers include

Instinct and Capacity	C. E. AYRES
Behavior and Intelligence	B. H. BODE
The Unity of Consciousness	MAURICE PICARD
Geneticism as a Heuristic Principle in Psychology	JOHN M. FLETCHER
Giving up Instincts in Psychology	ZING YANG KUO
The Intellectual Significance of the Grasping Reflex	R. M. GIVLER
A Tentative Analysis of the Primary Data of Psychology	J. R. KANTOR
Reconstruction in Mental Tests	BEARDSLEY RUML
The Sanity of Hamlet	TENNEY L. DAVIS
The Meaning of "Religion" and the Place of Mysticism in Religious Life	JAMES H. LEUBA
A Behavioristic View of Purpose	RALPH B. PERRY

Published on alternate Thursdays

SUB-STATION 84, NEW YORK CITY

$4 PER ANNUM, 26 NUMBERS 20 CENTS PER COPY

ARCHIVES OF PHILOSOPHY

Editorial communications should be addressed to

Professor FREDERICK J. E. WOODBRIDGE, **Columbia University,** New York City

Order should be sent to ARCHIVES OF PHILOSOPHY
Sub-Station 84, New York City

The numbers are as follows:

The Concept of Control. SAVILLA ALICE ELKUS. 40 cents.
The Will to Believe as a Basis for the Defense of Religious Faith. ETTIE STETTHEIMER. $1.
The Individual: A Metaphysical Inquiry. WILLIAM FORBES COOLEY. $1.00
The Ethical Implications of Bergson's Philosophy. UNA BARNARD SAIT. $1.25.
Religious Values and Intellectual Consistency. EDWARD H. REISNER. 75 cents.
Rosmini's Contribution to Ethical Philosophy. JOHN FAVATA BRUNO. 75 cents.
The Ethics of Euripides. RHYS CARPENTER. 50 cents.
The Logic of Bergson. GEORGE WILLIAM PECKHAM. 75 cents.
The Metaphysics of the Supernatural as illustrated by Descartes. LINA KAHN. $1.00.
Idea and Essence in the Philosophies of Hobbes and Spinoza. ALBERT G. A. BALZ. $1.25.
The Moral and Political Philosophy of John Locke. STERLING TOWER LAMPRECHT. $1.50.
Science and Social Progress. HERBERT WALLACE SCHNEIDER. $1.25.

DIRECTORY OF AMERICAN PSYCHOLOGICAL PERIODICALS

American Journal of Psychology—Ithaca, N. Y.: Morrill Hall.
Subscription $6.50. 600 pages annually. Edited by E. B. Titchener.
Quarterly. General and experimental psychology. Founded 1887.

Pedagogical Seminary—Worcester, Mass.: 959 Main Street.
Subscription $5. 575 pages annually. Edited by G. Stanley Hall.
Quarterly. Pedagogy and educational psychology. Founded 1891.

Psychological Review—Princeton, N. J.: Psychological Review Company.
Subscription $4.25. 480 pages annually.
Bi-monthly. General. Founded 1894. Edited by Howard C. Warren.

Psychological Bulletin. Princeton, N. J.: Psychological Review Company.
Subscription $5. 720 pages annually. Psychological literature.
Monthly. Founded 1904. Edited by Shepherd I. Franz.

Psychological Monographs. Princeton, N. J.: Psychological Review Company.
Subscription $5.50 per vol. 500 pp. Founded 1895. Ed. by James R. Angell.
Published without fixed dates, each issue one or more researches.

Psychological Index. Princeton, N. J.: Psychological Review Company.
Subscription $1.50. 200 pp. Founded 1895. Edited by Madison Bentley.
An annual bibliography of psychological literature.

Journal of Philosophy—New York: Sub-Station 84.
Subscription $4. 728 pages per volume. Founded 1904.
Bi-weekly. Edited by F. J. E. Woodbridge and Wendell T. Bush.

Archives of Psychology. Sub-Station 84, New York: Archives of Psychology.
Subscription $5. 500 pp annually. Founded 1906. Edited by R. S. Woodworth.
Published without fixed dates, each number a single experimental study.

Journal of Abnormal Psychology and Social Psychology—Boston.
Subscription $5. Richard G. Badger. Edited by Morton Prince.
Bi-Monthly. 432 pages annually. Founded 1906. Abnormal and social.

Psychological Clinic. Philadelphia: Psychological Clinic Press.
Subscription $2.50. 288 pages. Ed. by Lightner Witmer. Founded 1907.
Without fixed dates (9 numbers). Orthogenics, psychology, hygiene.

Training School Bulletin. Vineland, N. J.: The Training School.
Subscription $1. 160 pages annually. Edited by E. R. Johnstone. Founded 1904.
Monthly (10 numbers). Psychology and training of defectives.

Journal of Educational Psychology. Baltimore: Warwick & York.
Subscription $4. 540 pages annually. Founded 1910.
Monthly (9 numbers). Managing Editor, Harold O. Rugg.
(Educational Psychology Monographs.
Published separately at varying prices. Same publishers.)

Comparative Psychology Monographs—Baltimore: Williams & Wilkins Co.
Subscription $5. 500 pages per volume. Edited by W. S. Hunter.
Published without fixed dates, each number a single research.

Psychoanalytic Review. Washington, D. C.: 3617 10th Street, N. W.
Subscription $6. 500 pages annually. Psychoanalysis.
Quarterly. Founded 1913. Edited by W. A. White and S. E. Jelliffe.

Journal of Experimental Psychology—Princeton, N. J.
Psychological Review Company. 480 pages annually. Experimental.
Subscription $4.25. Founded 1916. Bi-monthly. Ed. by John B. Watson.

Journal of Applied Psychology. Worcester, Mass.: Florence Chandler.
Subscription $4. 400 pages annually. Founded 1917.
Quarterly. Edited by James P. Porter and William F. Book.

Journal of Comparative Psychology—Baltimore: Williams & Wilkins Co.
Subscription $5. 500 pages per volume. Founded 1921.
Bi-monthly. Edited by Knight Dunlap and Robert M. Yerkes.

INDIVIDUAL DIFFERENCES AS AFFECTED BY PRACTICE

BY

GEORGINA STICKLAND GATES, Ph. D.

Stanford Library

AUG 12 1923

ARCHIVES OF PSYCHOLOGY

EDITED BY R. S. WOODWORTH

No. 58

NEW YORK

AUGUST, 1922

AGENTS: G. E. STECHERT & CO. : London (2 Star Yard, Carey St., W. C.); Paris (16, rue de Condé)

ARCHIVES OF PSYCHOLOGY

SUB-STATION 84, NEW YORK CITY. The Subscription price is five dollars per volume of about 500 pages. Volume I comprises Nos. 1-10, Volume II, Nos. 11-18, Volume III, Nos. 19-25, Volume IV, Nos. 26-32, Volume V, Nos. 33-39, Volume VI, Nos. 40-46, Vol. VII, Nos. 47-52, and Vol. VIII, Nos. 53-58. The available numbers are as follows:

2. On the Functions of the Cerebrum: The Frontal Lobes: SHEPHERD IVORY FRANZ. 50c.
3. Empirical Studies in the Theory of Measurements: EDWARD L. THORNDIKE. 50c.
4. Rhythm as a Distinguishing Characteristic of Prose Style: ABRAM LIPSKY. 50c.
5. The Field of Distinct Vision: W. C. RUEDIGER. 70c.
6. The Influence of Bodily Position on Mental Activities: ELMER E. JONES. 50c.
7. A Statistical Study of Literary Merit: FREDERIC LYMAN WELLS. 80c.
8. The Relation between the Magnitude of the Stimulus and the Time of Reaction: SVEN FROEBERG. 35c.
9. The Perceptual Factors in Reading: FRANCIS MARION HAMILTON. 50c.
10. Time in English Verse Rhythm: WARNER BROWN. 70c.
11. The Hearing of Primitive Peoples: FRANK G. BRUNER. $1.00.
12. Studies in Development and Learning: EDWIN A. KIRKPATRICK. $1.00
13. The Inaccuracy of Movement: H. L. HOLLINGWORTH. 80c.
14. A Quantitative Study of Rhythm: HERBERT WOODROW. 60c.
16. On certain Electrical Processes in the Human Body and their Relation to Emotional Reactions: FREDERIC LYMAN WELLS and ALEXANDER FORBES. 40c.
17. The Relative Merit of Advertisements: EDWARD K. STRONG, JR. $1.00.
18. Attention and Movement in Reaction Time: J. V. BREITWIESER. 50c. (Cloth, 75c.)
19. An Empirical Study of Certain Tests for Individual Differences: MARY THEODORA WHITLEY. $1.25. (Cloth, $1.50.)
20. Visual Acuity with Lights of Different Colors and Intensities: DAVID EDGAR RICE. 50c. (Cloth, 75c.)
21. The Curve of Forgetting: C. H. BEAN. 45c. (Cloth, 70c.)
22. The Influence of Caffein on Mental and Motor Efficiency: H. L. HOLLINGWORTH. $1.50. (Cloth, $1.75.)
23. Reaction Time to Retinal Stimulation. A. T. POFFENBERGER, JR. 70c. (Cloth, 95c.)
24. Interference and Adaptability: ARTHUR JEROME CULLER. 75c. (Cloth, $1.00.)
25. Reaction to Multiple Stimuli: JOHN WELHOFF TODD. 60c. (Cloth, 85c.)
26. A Study in Incidental Memory: GARRY C. MYERS. $1.00. (Cloth, $1.25.)
27. A Statistical Study of Eminent Women: CORA SUTTON CASTLE. 80c. (Cloth, $1.05.)
28. The Mental Capacity of the American Negro: MARION J. MAYO. 60c. (Cloth, 85c.)
29. Experimental Studies in Judgment: H. L. HOLLINGWORTH. $1.25. (Cloth, $1.50.)
30. The Psychological Researches of James McKeen Cattell: A Review by Some of His Pupils. $1.00. (Cloth, $1.25.)
31. Fatigue and Its Effects upon Control: ISAAC EMERY ASH. 60c. (Cloth, 85c.)
32. The Transfer Effects of Practice in Cancellation Tests: MELVIN ALBERT MARTIN. 60c. (Cloth, 85c.)
33. The Intellectual Status of Children Who are Public Charges: J. L. STENQUIST, E. L. THORNDIKE, and M. R. TRABUE. 50c. (Cloth, 75c.)
34. The Relation of Quickness of Learning to Retentiveness: DARWIN OLIVER LYON. 50c. (Cloth, 75c.)
35. The Overcoming of Distraction and Other Resistances: JOHN J. B. MORGAN. 75c. (Cloth, $1.00.)
36. The Psychology of the Negro—An Experimental Study: GEORGE OSCAR FERGUSON, JR. $1.25. (Cloth, $1.50.)
37. The Effect of Distraction on Reaction Time: JOHN E. EVANS: $1.00. (Cloth, $1.25.)
38. The Effect of Humidity on Nervousness and on General Efficiency: LORLE IDA STECHER. 90c. (Cloth, $1.15.)
39. The Mechanism of Controlled Association: MARK A. MAY. 75c. (Cloth, $1.00.)
40. Recitation as a Factor in Memorizing: ARTHUR I. GATES. $1.00. (Cloth, $1.25.)
41. Mental Fatigue during Continuous Exercise of a Single Function: THOMAS RUSSELL GARTH. 85c. (Cloth, $1.10.)
42. A Psychological Study of Trade-Mark Infringement: RICHARD H. PATTERSON, JR. 85c. (Cloth, $1.10.)
43. Individual Differences and Family Resemblances in Animal Behavior: HALSEY J. BAGG. 70c. (Cloth, $1.00.)
44. Experimental Studies in Recall and Recognition: EDITH MULHALL ACHILLES. 90c. (Cloth, $1.25.)
45. The Morphologic Aspect of Intelligence: SANTE NACCARATI. 70c. (Cloth, $1.10.)
46. Psychological Examinations of College Students: F. EDITH CAROTHERS. $1.25.
47. The Effects of Practice on Judgments of Absolute Pitch: EVELYN GOUGH. $1.25
48. An Experimental Study of Silent Thinking: RUTH S. CLARK. $1.40.
49. Some Empirical Tests in Vocational Selection: HERBERT W. ROGERS. 75 cents.
50. Adenoids and Diseased Tonsils: Their Effect on General Intelligence. MARGARET COBB. $1.00.
51. An Experimental Study of the Factors and Types of Voluntary Choice. ALFRED H. MARTIN. $1.50.
52. Some Well-Known Mental Tests Evaluated and Compared. DOROTHY RUTH MORGENTHAU. 80 cents.
53. Mood in Relation to Performance. ELIZABETH T. SULLIVAN. $1.00.
54. The Influence of Incentive and Punishment upon Reaction Time: ALBERT M. JOHANSON. 80c.
55. Psychological Tests Applied to Factory Workers: EMILY THORP BURR. $1.25.
56. A Study of the Relation of Accuracy to Speed: HENRY E. GARRETT. $1.25.
57. An Experimental Study of Hunger in Its Relation to Activity: TOMI WADA. $1.50.
58. Individual Differences as Affected by Practice: GEORGINA STICKLAND GATES. $1.00.
59. Studies in Industrial Psychology: ELSIE OSCHRIN BREGMAN. 75 cents.

ARCHIVES OF PSYCHOLOGY

SUB-STATION 84, NEW YORK CITY

In addition to the numbers of the Archives, the following monographs are to be obtained from us:

Measurements of Twins: EDWARD L. THORNDIKE. 50 cents.
Avenarius and the Standpoint of Pure Experience: WENDELL T BUSH. 75 cents.
The Psychology of Association: FELIX ARNOLD. 50 cents.
The Measurement of Variable Quantities: FRANZ BOAS. 50 cents.
Linguistic Lapses: FREDERIC LYMAN WELLS. $1.00.
The Diurnal Course of Efficiency: HOWARD D. MARSH. 90 cents.
The Time of Perception as a Measure of Differences in Sensations: VIVIAN ALLEN CHARLES HENMON. 60 cents.
Interests in Relation to Intelligence: LOUISE E. POULL. $1.00. Reprinted from *Ungraded.*
The Conditioned Pupillary and Eyelid Reactions: HULSEY CASON. $1.00. Reprinted from the *Journal of Experimental Psychology.*

THE JOURNAL OF PHILOSOPHY

SUB-STATION 84, NEW YORK CITY

Published on alternate Thursdays

$4 PER ANNUM, 26 NUMBERS 20 CENTS PER COPY

Edited by Professors F. J. E. WOODBRIDGE and WENDELL T. BUSH of Columbia University.

There is no similar journal in the field of scientific philosophy. It is identified with no philosophical tradition, and stands preeminently for the correlation of philosophy with the problems and experience of the present.

Among the many psychological articles appearing in this Journal are the following of recent date:

Measures of Intelligence: A. T. POFFENBERGER
The Identity of Instinct and Habit: KNIGHT DUNLAP
Must We Give Up Instincts in Psychology? J. R. GEIGER
The Modification of Instinct: WALTER S. HUNTER
Mr. Russell's Psychology: F. C. S. SCHILLER
Intelligence and Intellect: A. A. ROBACK

ARCHIVES OF PHILOSOPHY

Editorial communications should be addressed to

Professor FREDERICK J. E. WOODBRIDGE, Columbia University, New York City

Order should be sent to ARCHIVES OF PHILOSOPHY Sub-Station 84, New York City

The numbers are as follows:

The Concept of Control. SAVILLA ALICE ELKUS. 40 cents.
The Will to Believe as a Basis for the Defense of Religious Faith. ETTIE STETTHEIMER. $1.
The Individual: A Metaphysical Inquiry. WILLIAM FORBES COOLEY. $1.00
The Ethical Implications of Bergson's Philosophy. UNA BARNARD SAIT. $1.25.
Religious Values and Intellectual Consistency. EDWARD H. REISNER. 75 cents.
Rosmini's Contribution to Ethical Philosophy. JOHN FAVATA BRUNO. 75 cents.
The Ethics of Euripides. RHYS CARPENTER. 50 cents.
The Logic of Bergson. GEORGE WILLIAM PECKHAM. 75 cents.
The Metaphysics of the Supernatural as illustrated by Descartes. LINA KAHN. $1.00.
Idea and Essence in the Philosophies of Hobbes and Spinoza. ALBERT G. A. BALZ. $1.25.
The Moral and Political Philosophy of John Locke. STERLING TOWER LAMPRECHT. $1.50.
Science and Social Progress. HERBERT WALLACE SCHNEIDER. $1.25.